ETSI TECHNICAL REPORT

ETR 250

July 1996

Source: ETSI TC-TM

Reference: DTR/TM-05006

ICS: 33.040.20

Key words: transmission, network, 3,1 kHz, speech, quality, planning, methodology

Transmission and Multiplexing (TM); Speech communication quality from mouth to ear for 3,1 kHz handset telephony across networks

ETSI

European Telecommunications Standards Institute

ETSI Secretariat

Postal address: F-06921 Sophia Antipolis CEDEX - FRANCE
Office address: 650 Route des Lucioles - Sophia Antipolis - Valbonne - FRANCE
X.400: c=fr, a=atlas, p=etsi, s=secretariat - **Internet:** secretariat@etsi.fr

Tel.: +33 92 94 42 00 - Fax: +33 93 65 47 16

Contents

Blank page

Foreword

This ETSI Technical Report (ETR) has been produced by Transmission and Multiplexing (TM) Technical Committee of the European Telecommunications Standards Institute (ETSI).

ETRs are informative documents resulting from ETSI studies which are not appropriate for European Telecommunication Standard (ETS) or Interim European Telecommunication Standard (I-ETS) status. An ETR may be used to publish material which is either of an informative nature, relating to the use or the application of ETSs or I-ETSs, or which is immature and not yet suitable for formal adoption as an ETS or an I-ETS.

Introduction

It is important to achieve a high probability that users are provided with adequate transmission quality for voice telephony. The traditional transmission planning methods are subject to change because of:

a) the digitalisation of networks which makes delay and echo more critical than loudness;

b) the changes in regulations, in many countries, which mean that connections may be provided by a combination of different networks, managed by several different operators;

c) the growth in new mobile networks which can provide services in circumstances where service was not previously possible, but where the quality may be affected by the need to use low bit-rate coding to conserve the radio spectrum;

d) the growth in the use of devices where the output is not linearly related to the input, such as low bit-rate codecs and noise reduction systems;

e) the increased use of Digital Circuit Multiplication Equipment (DCME) and Voice Packeting Equipment (VPE);

f) the introduction of new technology such as Asynchronous Transfer Mode (ATM), flexible routing schemes, etc.

In these changing circumstances, network operators need guidance and a satisfactory method for assessing the transmission quality of connections and predicting whether or not the user will be satisfied. This method needs to be computational and based on planning values, because it is not possible for measurements to be made to check the performance of all possible paths through real networks. The method needs also to take account of the additional utility of, for example, mobile networks, where users may be prepared to accept a lower voice transmission quality relative to what is usually achieved in wireline networks, if thereby they can have a service where previously none was available.

Transmission planning in the past has tended to evaluate transmission parameters individually, often on a worst case basis. The effects of the combination of different parameters and the availability of mobile services have often not been considered due to a lack of information.

This ETR addresses the mouth to ear performance of nominal 3,1 kHz voice telephony via handsets across networks of all types. It may be extended later to cover non-handset telephony and wideband (7 kHz) telephony.

NOTE: Voice transmission quality is only one aspect of the "usability" or "utility" of telecommunication services which is under discussion in International Standards Organization (ISO) and ETSI. The usability concept includes the two dimensions *performance (of the system)* and *attitude (of the users).* The voice transmission quality affects both dimensions, a fact which is recognised in the computation model, at least on a provisional basis. (Price considerations is of course also an important "utility" factor which needs assessment. However, adequate data is not available for inclusion of a "price factor" in the model). For further information, see annex J.

Blank page

1 Scope

This ETR applies to mouth to ear narrowband (nominally 3,1 kHz) telephony connections via handsets across one or more telecommunications networks of any type, public or private, fixed or mobile.

The ETR provides:

a) a compendium of established and well recognised transmission planning information on individual parameters, including limits for satisfactory performance and information on equipment with special transmission characteristics;

b) a new simple computational planning method to consider the combination effect of transmission parameters when evaluating transmission impairments for telephony services, including estimation of users' opinion using the scales percentage poor or worse, percentage good or better, percentage terminating early and mean opinion score.

The computational planning method is an adaptation of computation models published in ITU-T (former CCITT) documentation, complemented with the result from some recently presented results of subjective tests.

The objective is to assist network operators in ensuring that users will be satisfied with the transmission performance whilst avoiding over-engineering of the networks. The aim is to give realistic, practical guidance rather than a scientifically exact treatment of quality factors.

It should be noted that the computational planning method, which is described in clause 9, has not yet been fully verified and may give incorrect results. Until this verification has been completed, predictions resulting from the use of this planning method should therefore be treated with caution.

2 References

This ETR incorporates by dated and undated reference, provisions from other publications. These references are cited at the appropriate places in the text and the publications are listed hereafter. For dated references, subsequent amendments to or revisions of any of these publications apply to this ETR only when incorporated in it by amendment or revision. For undated references the latest edition of the publication referred to applies.

NOTE: With regard to Recommendations from ITU-T (former CCITT), the following principle has been followed: The designation "CCITT" is used for Recommendations published before and including the Blue Books (1988). Recommendations (new or revised) after 1989 are designated "ITU-T".

[1] A. Coleman, N. Gleiss, J. Sotscheck, H. Scheuermann, P. Usai: "Subjective Performance Evaluation of the RPE-LTP Codec for the Pan-European Cellular Digital Mobile Radio System", IEEE Proceedings GLOBECOM ‘89, Dallas, Texas.

[2] C. South, J. Rugelback, P. Usai, N. Kitawaki, J. Rosenberger, J.R. Cavanaugh, C.A. Adessanya, D. Pascal, N. Gleiss, G.J. Barnes: "Subjective Performance Assessment of CCITT´s 16 kbit/s Speech Coding Algorithm", Speech Communication 12, 1993, No.2.

[3] C. South, P. Usai "Subjective Performance of CCITT´s 16 kbit/s LD-CELP Algorithm with Voice Signals", IEEE Proceedings GLOBECOM ´92, Orlando, Florida.

[4] CCITT COM 12 - D.18: "Considerations of Low Bit-Rate Codecs in Transmission Planning", Ellemtel, Feb.1994.

[5] J. Alnatt: "Subjective Rating and Apparent Magnitude"; Int. J. Man - Machine Studies (1975) 7.

[6] CCITT COM 15-33: "Technical Report on Echo Cancellers", AT&T, July 1993.

[7] CCITT COM 15/1-TD 7: "Report of the Rapporteur for very Low Bit-Rate Visual Telephony", 6.-17. Sept.1993.

[8] CCITT COM XII-21: "The Effect of Delayed Sidetone on the Overall Sound Quality of a Telephone Connection", Swedish Telecom 1989-1992.

[9] CCITT COM XII-25: "Effect of Talker Echo on the Quality of Telephone Connections. Results from a conversational experiment", France 1989-1992.

[10] CCITT COM XII-No. 158-E, June 1983; "Subjective Effects of Sidetone".

[11] CCITT COM XV-D.465: "Evaluation of LD-CELP Performance for Use in DCME", COMSAT, INTELSAT, Nov 1992.

[12] CCITT Recommendation G.105 (1988): "Hypothetical reference connection for crosstalk studies".

[13] CCITT Recommendation G.123 (1988): "Circuit noise in national networks".

[14] CCITT Recommendation G.131 (1988): "Stability and echo".

[15] CCITT Recommendation G.134 (1988): "Linear crosstalk".

[16] CCITT Recommendation G.151 (1988): "General performance objectives applicable to all modern international circuits and national extension circuits".

[17] CCITT Recommendation G.153 (1988): "Characteristics appropriate to international circuits more than 2 500 km in length".

[18] CCITT Recommendation G.164 (1988): "Echo suppressors".

[19] ITU-T Recommendation G.165 (1993): "Echo cancellers".

[20] CCITT Recommendation G.222 (1988): "Noise objectives for design of carrier-transmission systems of 2500 km".

[21] CCITT Recommendation G.223 (1988): "Assumptions for the calculation of noise on hypothetical reference circuits for telephony".

[22] CCITT Recommendation G.226 (1988): "Noise on a real link".

[23] CCITT Recommendation G.228 (1988): "Measurement of circuit noise in cable systems using a unifor-spectrum random noise loading".

[24] ITU-T Recommendation G.712 (1992): "Transmission performance characteristics of pulse code modulation".

[25] ITU-T Recommendation G122 (1993): "Influence of national systems on stability, talker echo in international connections".

[26] ITU-T Recommendation P.11 (1993): "Effect of transmission impairments".

[27] CCITT Recommendation P.16 (1988): "Subjective effects of direct crosstalk; tresholds of audibility and intelligibility".

[28] ITU-T Recommendation P.65 (1993): "Objective instrumentation for the determination of loudness ratings".

[29] ITU-T Recommendation Q.551 (1994): "Transmission characteristics of digital exchanges".

[30] ITU-T Recommendation Q.552 (1994): "Transmission characteristics at 2-wire analogue interfaces of digital exchanges".

[31] ITU-T Recommendation Q.553 (1994): "Transmission characteristics at 4-wire analogue interfaces of digital exchanges".

[32] ITU-T Recommendation O.41 (1993): "Psophometer for use on telephone-type circuits".

[33] ITU-T Recommendation P.79 (1993): "Calculation of loudness ratings for telephone sets".

[34] ITU-T Recommendation G.111 (1993): "Loudness ratings (LRs) in an international connection".

[35] ITU-T Recommendation G.113 (1993): "Transmission impairments".

[36] ITU-T Recommendation G.114 (1993): "One-way transmission time".

[37] ITU-T Recommendation G.173 (1993): "Transmission planing aspects of the speech service in digital public land mobile networks".

[38] ITU-T Recommendation G.121 (1993): "Loudness ratings (LRs) of national systems".

[39] ITU-T Recommendation P.31 (1993): "Transmission characteristics for digital telephones".

[40] ITU-T Recommendation P.64 (1993): "Determination of sensitivity/frequency characteristics of local telephone systems".

[41] CCITT Revision Supplement No. 11, COM XII-R30-E, April 1992; "Some Effects of Sidetone".

[42] ITU-T Recommendation Q.115 (1993): "Control of echo suppressors - Control of echo suppressors and echo cancellers".

[43] CCITT Supplement No. 3 page 248-329 (1988): "Models for Predicting Transmission Quality from objective Measurements".

[44] CCITT Supplement No. 13 to Sections 2 to 5 of the Series G Recommendations (1988; "Noise at the Terminals at the Battery Supply").

[45] TBR 8 (1994): "ISDN, Telephony 3.1 kHz Teleservice, Attachment Requirements for Handset Terminals".

[46] ITU-T Recommendation Q.2764 (1995): "Broadband integrated services digital network (B-ISDN) - Signalling system No. 7 B-ISDN user part (B-ISUP) - Basic call procedures".

[47] ETS 300 175-8 (1992): "Radio Equipment and Systems (RES); Digital European Cordless Telecommunications (DECT) Common interface Part 8: Speech coding and transmission".

[48] G. Williams, H. Suyderoud "Subjective Performance Evaluation of the 32 kbit/s ADPCM Algorithm", IEEE Proceedings GLOBECOM ´84, Vol.2, pp. 778-785.

[49] G. Modena, A. Coleman, P. Coverdale, P. Usai "Subjective Performance Evaluation of the 64 kbit/s (7 kHz) Audio Coder", IEEE Proceedings GLOBECOM ´86, Houston, Texas.

[50] "Handbook on Telephonometry" (Geneva 1987).

[51] ITU-T Recommendation V.34 (1994): "A modem operating at data signalling rates of up to 28 800 bits/s for use on the general switched telephone network and on leased point-to-point 2-wire telephone-type circuits".

[52] I-ETS 300 131 (1994): "Radio Equipment and Systems (RES); Common air interface specification to be used for the interworking between cordless telephone apparatus in the frequency band 864,1 MHz to 868,1 MHz, including public access services".

[53] J. Gruber, G. Williams: "Transmission Performance of Evolving Telecommunications Networks", (Artech House 1992).

[54] J. Fennick "Quality Measures and the Design of Telecommunications Systems", Artech House, ISBN 0-89006-258-7, 1988.

[55] J.R. Cavanaugh "A Model for the subjective Effects of Quantizing Distortion in Digital Transmission of Speech", Int. Conf. on Communications, June 1980.

[56] M.H. Sherif, D.O. Bowker, G. Bertocci, B.A. Orfort, G.A. Mariano: "Overview and Performance of CCITT/ANSI embedded ADPCM Algorithm", IEEE Trans. on Comm. Vol 41, No.2, Febr. 1993.

[57] N. Osaka, K. Kakehi, S. Lai, N. Kitawaki: "A Model for Evaluating Talker Echo and Sidetone in a Telephone Transmission Network.", IEEE Trans. on Comm. Vol 40, Nov. 1992.

[58] S. Furui, M.M. Sondhi "Advances in Speech Signal Processing", M. Dekker Inc., New York, 1991.

[59] S.R. Quackenbush, T.P. Barnwell III, M.A. Clemens "Objective Measures of Speech Quality", Prntice Hall, ISBN 0-13-629056-6, 1988.

[60] A. Gersho "Advances in Speech and Audio Compression", Proc. of the IEEE, Vol. 82, No. 6, June 1994.

[61] B.S. Atal, V. Cuperman and A. Gersho "Advances in Speech Coding", Norwell, MA, Kluwer, 1991.

[62] CCITT Recommendation G.711 (1988): "pulse code modulation (PCM) of voice frequencies".

[63] CCITT Recommendation G.721 (1988): "32 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)".

[64] CCITT Recommendation G.722 (1988): "7 kHz audio-coding within 64 kbit/s".

[65] CCITT Recommendation G.723 (1988): "Extensions of Recommendation G.721 ADPCM to 24 and 40 kbit/s for DCME Application".

[66] ITU-T Recommendation G.727 (1990): "5-, 4-, 3- and 2-bits/sample embedded adaptive differential pulse code modulation (ADPCM)".

[67] CCITT Recommendation P.82 (1988): "Method for evaluating of service from the standpoint of speech transmission quality".

[68] CCITT SG XII SQEG "Methods for Subjective Determination of Transmission Quality of Low Bit-rate Codecs", Rev. Recommendation P.80 & P.83, 1992.

[69] IS 54: "American standard for Digital Advanced Mobile Phone System (DAMPS)".

[70] COM 15-20: "Transmission Quality of interconnected PSTN-Digital Cellular Networks" (COMSAT).

[71] COM 15/1-TD 7: "Report of the Rapporteur for Very Low Bit-rate Visual Telephony", Sept. 1993.

[72] Contribution T1A1.6/92 060 from Tellabs, "Test Results of Modem/Echo Canceller Interaction", March 1992.

[73] D.G. Messerschmitt: "Echo Cancellation in Speech and Data Transmission", IEEE J. Selected Areas in Communications, Vol SAC-2, No.2, March 1984.

[74] "IEEE Recommended Practice for Speech Quality Measurements", IEEE Trans. on Audio and Electroacoustics, Sept. 1969.

[75] E. Kessler, A. Farquharson: "Transmission Quality of Digital Circuit Multiplication and Low Rate Encoding", British Telecommunications Engineering, Vol. 12, April 1993.

[76] E.M. Buttler, T.H. Curtis and H.G. Hagen: "An Analysis of Echo Canceller/Voiceband Data Set Interaction - Results of a Laboratory Simulation", Bell Laboratories Technical Memorandum, April 6, 1979. Available as T1A1 document T1A1.6/93-03, February 1993.

[77] G.S. Fang: "Voice Channel Echo Cancellation", IEEE Communications Magazine, December 1983.

[78] H.J.M. Steeneken and T. Houtgast: "A physical Method for Measuring Speech-Transmission Quality", Journal of the Acoustical Society of America, 87(1), Jan. 1980.

[79] I. Boyd: "Speech Coding for Telecommunications", Electronics & Communications Engineering Journal, October 1992.

[80] IEC 849 (1989): "Sound Systems for Emergency Purposes".

[81] ITU-T Recommendation G.728 (1992) "Coding of speech at 16 kbit/s using low-delay code excited linear prediction".

[82] ITU-T Recommendation G.763 (1993): "Digital circuit multiplication equipment using 32 kbit/s ADPCM and digital speech interpolation".

[83] ITU-T Recommendation G.764 (1990): "Voice packetization - packetized voice protocols".

[84] ITU-T Recommendation G.765 (1992): "Packet circuit multiplication equipment".

[85] ITU-T Recommendation P.80 (1993): "Methods for subjective determination of transmission quality".

[86] ITU-T Recommendation P.83 (1993): "Subjective performance assessment of telephone-band and wideband digital codecs".

[87] ITU-T Recommendation P.84 (1993): "Subjective listening test method for evaluationg digital circuit multiplication and packetized voice systems".

[88] M. Givens, J. D. Mills, M. H. Sherif, Y. T. Tao, J. A. Zebarth: "Draft Technical Report on Echo Cancellers in Digital Networks, version 1.3", T1A1.6 Editorial Group, July 13, 1992.

[89] John L. Sullivan:" Is Transmission Satisfactory ?", Bell Labs Record Vol.52, No. 3, March 1974.

[90] John S. Oakland: "Total Quality Management", Butterworth - Heinemann Ltd., Oxford 1993.

[91] K. Adler, G. Rahmig: "Unterdrückung des Sprecherechos in gemischt analog/digitalen Netzen", Frequenz, Band 46, Juli-August 1992.

[92] M.J. Miller, B. Vucetic, Les Berry: "SATELLITE COMMUNICATIONS Mobile and Fixed Services" Kluwer Academic Publishers, 1993.

[93] M.M. Sondhi and W. Kellerman: "Adaptive Echo Cancellation for Speech Signals, Advances in Speech Signal Processing", S. Furui and M.M. Sondhi, eds., Marcel Dekker, Inc., New York, 1992.

[94] N. Kitawaki: "Research of Objective Speech Quality Assessment", NTT Rev. Vol. 3, No. 5, Sept.1991.

[95] N. S. Jayant and P. Noll: "Digital Coding of Waveforms: Principles and Applications to Speech and Video", Prentice Hall, 1984.

[96] Norman Gleiss: "Usability - Concepts and Evaluation", TELE 2/92, English Edition.

[97] R. Loiola, A. Striuli (CSELT, Italy): "Asymmetry in Connection Quality", ETSI VTQM-E, Turin, June 1993, TD 93.34.

[98] R.E. Collins: "Service Quality - Measuring the Network and the Customer", Telephony Vol. 209, No. 24, Dec. 1985.

[99] T.E. Curtis, S.J. D'Ambra, R.H. Tegahoff and L.E. Ashkenazi: "Use of Digital Echo Canceller in the AT&T DOMSAT Intertoll Network", Proceedings of The Fifth International Conference on Digital Satellite Communications, March 1981, Genoa, Italy.

[100] Y.G. Tao, K.D. Kolwicz, C.W.K. Gritton and D.L. Duttweiler: "A cascadable VLSI Echo Canceller", J. Selected Areas in Communications, Vol SAC-2, No.2, March 1984.

[101] S. Wong: "An evaluation of 6.4 kbit/s speech codecs for INMARSAT-M system", ICASSP '91, Paper S9.16.

[102] Report on ITU-T SG12 Question 13/12: "Methods for measuring and modelling the effects of non-linear processes on the speech quality of transmission systems", Rapporteurs N. Kitawaki (Japan) & D.J. Atkinson (USA), Feb. 1994.

[103] Norman Gleiss: "Desirable Sending Frequency Response of Telephone Sets", Tele 1/89, English edition.

[104] J. D. Healy, J. M. Kelly, D. G. Leeper, M. G. Linnell and J. A. Maher: "How's Service? Ask the Customer", Bell Labs Record Vol. 60, Febr. 1982.

[105] ANSI T1.221 (1991): "In-service, Non intrusive Measurement Device (INMD) - Voice Service Measurements".

[106] ITU-T Draft Recommendation P.INMD: "In-service, non intrusive measurement device".

[107] ITU-T Recommendation P.50 (1993): "Artificial voices".

[108] D. L. Richards: "Telecommunication by Speech", Butterworth, ISBN-0-408-70344-X.

[109] ITU-T Recommendation P.34 (1993): "Transmission characteristics of hands-free telephones".

[110] CCITT Recommendation P.53 (1988): "Psophometers (apparatus for the objective measurement of circuit noise)".

[111] CCITT Recommendation P.55 (1988): "Apparatus for the measurement of impulsive noise".

[112] ITU-T Recommendation P.56 (1993): "Objective measurement of active speech level".

[113] ITU-T Recommendation P.81 (1993): "Modulated noise reference unit (MNRU)".

[114] ITU-T Recommendation G.100 (1993): "Definitions used in Recommendations on general characteristics of international telephone connections and circuits".

[115] ITU-T Recommendation G.101 (1993): "The transmission plan".

[116] ITU-T Recommendation G.126 (1993): "Listener echo in telephone networks".

[117] CCITT Recommendation G.143 (1988): "Circuit noise and the use of compandors".

[118] CCITT Recommendation G.161 (1988): "Echo suppressors suitable for circuits having either short or long propagation times".

[119] ITU-T Draft Recommendation G.174 (1994): "Transmission performance objectives for terrestrial digital wireless systems using portable terminals to access the PSTN".

[120] CCITT Recommendation G.229 (1988): "Unwanted modulation and phase jitter".

[121] ITU-T Recommendation G.726 (1990): "40, 32, 24, 16 kbit/s adaptive differential pulse code modulation (ADPCM)".

[122] CCITT Recommendation G.733 (1988): "Characteristics of primary PCM multiplex equipment operating at 1544 kbit/s".

[123] CCITT Recommendation G.792 (1988): "Characteristics common to all transmultiplexing equipment".

[124] ITU-T Recommendation G.766 (1992): "Facsimile demodulation/remodulation for digital circuit multiplication equipment".

[125] CCITT Recommendation E.125 (1988): "Inquiries among users of the international telephone service".

[126] CCITT COM 12-27: "Measurements of speech levels", Telia, December 1993.

[127] ETS 300 581 (1995): "European digital cellular telecommunications system; Half rate speech", (8 part ETS).

[128] I-ETS 300 036 (1992): "European digital cellular telecommunications system (phase 1); Full-rate speech transcoding (GSM 06.10)".

[129] I-ETS 300 037 (1992): "European digital cellular telecommunications system (phase 1); Substitution and muting of lost frames for full-rate speech traffic channels (GSM 06.11)".

[130] I-ETS 300 038 (1992): "European digital cellular telecommunications system (phase 1); Comfort noise aspects for full-rate speech traffic channels (GSM 06.12)".

[131] I-ETS 300 039 (1992): "European digital cellular telecommunications system (phase 1); Discontinuous transmission (DTX) for full-rate speech traffic channels (GSM 06.31)".

[132] I-ETS 300 040 (1992): "European digital cellular telecommunications system (phase 1); Voice activity detection (GSM 06.32)".

[133] I-ETS 300 176 (1992): "Radio Equipment and Systems (RES); Digital European European Cordless Telecommunications (DECT) Approval test specification".

[134] TBR 9: "European digital cellular telecommunications system; Attachment requirements for Global System for Mobile communication (GSM) mobile stations; Telephony".

[135] TBR 10: "Radio Equipment and Systems (RES); Digital European Cordless Telecommunications (DECT); General terminal attachment requirements; Telephony applications".

[136] I-ETS 300 020 (1992) Parts 1 and 2: "European digital cellular telecommunications system (phase 1); Mobile station conformity specifications (GSM 11.10)".

[137] ETS 300 326 (1993) Parts 1, 2 and 3: "Radio Equipment and Systems (RES); Terrestrial Flight Telephone System (TFTS)".

[138] ITU-T Recommendation X.25 (1993): "Interface between Data Terminal Equipment (DTE) and Data Circuit-terminating Equipment (DCE) for terminals operating in the packet mode and connected to public data networks by dedicated circuits".

[139] ITU-T Recommendation T.30 (1993): "Procedure for document facsimile transmission in the general switched telephone network".

[140] ITU-T Recommendation V.21 (1994): "300 bits per second duplex modem standardized for use in the general switched telephone network".

[141] CCITT Recommendation V.23 (1988): "600/1200-baud modem standardized for use in the general switched telephone network".

[142] CCITT Recommendation V.26 (1988): "2400 bits per second modem standardized for use on 4-wire leased telephone-type circuits".

[143] CCITT Recommendation V.27 (1988): "4800 bits per second modem with manual equalizer standardized for use on leased telephone-type circuits".

[144] CCITT Recommendation V.27ter (1988): "4800/2400 bits per second modem standardized for use in the general switched telephone network".

[145] CCITT Recommendation V.29 (1988): "9600 bits/s modem standardized for use on point-to-point 4-wire leased telephone-type circuits".

[146] ITU-T Recommendation V.32 (1993): "A family of 2-wire, duplex modems operating at data signalling rates of up to 9600 bit/s for use on the general switched telephone network and on leased telephone-type circuits".

[147] ISO 8402 (1994): "Quality management and quality assurance - Vocabulary".

[148] ITU-T Recommendation Q.764 (1993): "Signalling system No. 7 - ISDN user part signalling procedures".

[149] ITU-T Recommendation Q.767 (1991): "Application of the ISDN user part of CCITT signalling system No. 7 for international ISDN interconnections".

[150] ISO Draft CD 9241-11: "Guidance on usability specification and measures", 2 May 1992.

[151] ISO 226 (1987): "Acoustics - Normal equal-loudness level contours".

[152] IEC Publication 179: "Precision sound level meters".

[153] ANSI T1.508 (1992): "Network performance loss plan for evolving digital networks".

[154] ANSI T1.303 (1989): "Telecommunications - Digital processing of voice-band signals - Algorithms for 24, 32 and 40 kbit/s adaptive differential pulse-code modulation (ADPCM)".

[155] CCITT Fascile I.3 (1988): "Terms and definitions; abbreviations and acronyms".

[156] US Department of Defence Federal Standard 1016: "Telecommunications: Analog to digital conversion of radio voice by 4,800 bit/second code excited linear prediction (CELP)".

3 Definitions and abbreviations

General definitions are the same as used in ITU-T. For reference, see CCITT Fascicle I.3, "Terms and definitions; abbreviations and acronyms" [155].

Definitions and abbreviations that are special to this ETR appear at the appropriate places in the text, except for some definitions, which are given here.

3.1 Definitions

For the purposes of this ETR, the following definitions apply:

voice transmission quality: A measure of fundamental speech quality derived subjectively by direct comparison with a "trunk quality" communications system.

expectation factor: A normally positive quantity that represents the advantage of access that certain systems have over conventional wirebound handset telephony. It can represent parameters such as mobility, loudspeaking and cost which may be considered by a customer as advantages to offset related perceived degradations. An expectation factor might be negative, where for example a higher than normal voice transmission quality might be expected (e.g. wirebound telephony).

speech communication quality: A user's perception of voice transmission quality modified by his expactations of the communications system performance and its application, with special regard to his particular needs.

trunk (in US "toll") quality: A quality characterised by good intelligibility, good speaker identification, naturalness and with only minor disturbing impairments, for example, the average speech quality of long distance Public Switched Telephone Network (PSTN) connections.

communications quality: A quality characterised by good intelligibility, speaker indentity maintained, but with some loss of quality when directly compared with trunk quality, for example, the speech quality present in many mobile communications systems.

3.2 Abbreviations

For the purposes of this ETR, the following abbreviations apply:

AAL	ATM Adaptation Layer
ACR	Absolute Category Rating
ADPCM	Adaptive Differential Pulse Code Modulation
ANCOVA	ANalysis of COVAriance
ANOVA	ANalysis Of VAriance
ATC	Adaptive Transform Coding
ATM	Asynchronous Transfer Mode
B-ISDN	Broadband Integrated Services Digital Network
BcTR	Bellcore Transmission Rating
CBR	Constant Bit-Rate
CCITT	The International Telegraph and Telephone Consultative Committee
CED	CallED station identification
CELP	Code Excited Linear Prediction
CLR	Circuit Loudness Rating
CPE	Customer Premises Equipment
CRE	Corrected Reference Equivalent
DCME	Digital Circuit Multiplication Equipment
DCR	Degration Category Rating
DCS	Digital Communications System
DECT	Digital European Cordless Telecommunications
DOD	US Department Of Defence
DS1	Digital Standard 1
DSI	Digital Speech Interpolation
DTDT	Double-Talk Detection Threshold
DTMF	Dual Tone Multi-Frequency
DTX	Discontinuous Transmission
e.m.f.	electromotive force
EARS	Electro-Acoustic Reference System
EC	Echo Canceller
ECD	Echo Control Device
EL	Echo Loss
equ.	equation
ERP	Ear Reference Point
ES	Echo Suppressor
FDM	Frequency Division Multiplex
FPLMTS	Future Public Land Mobile Telecommunications System
GOB	Good Or Better
GSM	Global System for Mobile communications
HF	Human Factors
I-ETS	Interim European Telecommunications Standard
ICP	International Connecting Point
IEEE	The Institute of Electrical and Electronics Engineers
IMBE	Improved Multi Band Excitation
INMD	In-service, Non-intrusive Measuring Device
IRS	Intermediate Reference System
ISC	International Switching Center
ISDN	Integrated Services Digital Network
ISO	Internation Standardization Organisation
ITU	International Telecommunication Union
ITU-R	ITU Radio Standardization Sector (formerly CCIR)
ITU-T	ITU Telecommunication Standardization Sector (former CCITT)
LD-CELP	Low Delay Codebook Excited Linear Prediction
LEL	Listener Echo Loss
LELR	Listener Echo Loudness Rating
LPC	Linear Predictive Coding
LRE	Low Rate Encoding
LSTR	Listener SideTone masking Rating
MAF	Minimum Audible Field
MH	Murray Hill

MIPS	Million Instructions Per Second
MNRU	Modulated Noise Reference Unit
MOPS	Million Operations Per Second
MOS	Mean Opinion Score
MPE	Multi-Pulse Excited
MPEG	Moving Pictures Experts Group
MPLPC	Multi-Pulse Linear Predictive Coding
MRP	Mouth Reference Point
MTF	Modulation Transfer Function
NEST	Near-End Speech Threshold
NLP	Non-Linear Processor
OLR	Overall Loudness Rating
OLL	Open Loop Loss
OOLR	Overall Objective Loudness Rating
OREM	Objective Reference Equivalent Measurement
PBX	Private Branch Exchange
PCM	Pulse Code Modulation
PCME	Packet Circuit Multiplication Equipment
PDN	Packet Data Network
PLMN	Public Land Mobile Network
POW	Poor Or Worse
PSTN	Public Switched Telephone Network
qdu	quantizing distortion unit
r.m.s.	root mean square
RCRE	Receive Corrected Reference Equivalent
RLR	Receive Loudness Rating
ROLR	Receiver Objective Loudness Rating
RPE-LTP	Residual Pulse Exitation - Long Term Predictor
S/N	Signal to Noise
SBC	Sub-Band Coding
SCRE	Send Corrected Reference Equivalent
SD	Spectral Distance
SDH	Synchronous Digital Hierarchy
SLR	Send Loudness Rating
SNR	Signal-to quantization Noise Ratio
SQEG	Speech Quality Experts Group
SS7	Signalling System No. 7
St.D.	Standard Deviation
STC	Sub-Technical Committee
STI	Speech Transmission Index
STMR	SideTone Masking Rating
STRE	SideTone Reference Equivalent
TC	Technical Committee
TCH-HS	Traffic CHannel Half Size
TCL	Terminal Coupling Loss
TCLw	Weighted Terminal Coupling Loss
TDM	Time Division Multiplex
TELR	Talker Echo Loudness Rating
TETRA	Trans-European Trunked RAdio
TFTS	Terrestrial Flight Telecommunications System
TME	Terminate Early
TMN	Telecommunication Management Network
TOLR	Talker Objective Loudness Rating
TQI	Transmission Quality Index
TRP	Transmission Reference Point
UMTS	Universal Mobile Telecommunications System
VICP	Virtual International Connecting Point
VPE	Voice Packeting Equipment
VSELP	Vector Sum Excited Linear Prediction
WEPL	Weighted Echo Path Loss
XRLR	Crosstalk Receive Loudness Rating

4 Typical network components and configurations

With respect to this ETR, the communication channel between Mouth and Ear of a speech conversation consists of several different components forming the connection configuration. When considering the resulting overall transmission quality during network planning, it is necessary to be familiar with these components, the possible network configurations and their influence to the transmission quality. The following clauses of this ETR describe in more detail these different sources of impairments.

4.1 Network components

All the components forming a connection can be divided into three main groups: Terminal Elements, Connection Elements and Transmission Elements.

Terminal Elements with respect to speech transmission are all types of digital or analogue telephones, including cordless or mobile telephone sets, and includes the acoustic interface to the human mouth and ear. These components, with their Send Loudness Rating (SLR) and Receive Loudness Rating (RLR), contribute to the Overall Loudness Rating (OLR) of a connection. Further impairments can be caused by the Sidetone Masking Rating (STMR), the Listener Sidetone (LSTR) in conjunction with the D-Factor (dependent on the design of the handset), the frequency response in send and receive direction and some noise floor. In case of wireless systems, additional distortions will be added according to the used coding and modulation algorithm for the radio interface.

Connection Elements are all types of switching equipment, such as local exchanges, toll exchanges in public networks and PBXs in private networks. They use either an analogue or a digital switching matrix. Analogue systems may mainly contribute by noise, loss and in case of 4-wire to 2-wire conversion, with signal reflections as a source for echo effects. Digital exchanges will contribute mainly with transmission time due to signal processing.

Transmission Elements are all kinds of media, used as a link between connection elements and to the terminal elements. The physical media of these elements can be metallic, fiber-optic or wireless. The signal form is either analogue or digital.

In case of analogue signal transmission, cables contribute by loss and a shape in the frequency response. An additional impairment can be caused by noise due to longitudinal interference. Using the digital signal form, the main transmission impairment is caused by the propagation time via metallic and optical media. Wireless sections primarily contribute with a high value of delay in case of satellite links.

Beside the transmission of only one single communication channel via the different media, multiplexing is also used to transport several channels via one single physical medium. A variety of multiplexing systems are in use in existing networks. Analogue networks have been equipped with Frequency Division Multiplexing systems (FDM), mainly contributing with noise, transmission time and group delay distortion. Time Division Multiplexing systems (TDM) use Pulse Code Modulation (PCM) to transport several channels on one single medium. A major influence to transmission quality of these systems is caused by the transmission time. Modern equipment uses special coding algorithms to reduce the bit-rate of each communication channel. An example of such a device is an equipment called Digital Circuit Multiplication Equipment (DCME). Those systems contribute additional impairments e.g. quantizing noise and delay. Another source for delay arises in conjunction with the use of packetization techniques e.g. Asynchronous Transfer Mode (ATM) within the networks.

4.2 Network configurations

Beside the impairments caused by the different components, the main influence to the voice transmission quality is dependent on the specific network configuration. The configuration is formed by a tandeming of different network components, to connect two terminal elements in a communication channel, resulting in an addition of all impairments caused by every component.

The variety of possible network configurations is too extensive to show all of them within this ETR. Network configurations depend on the type of connection - short, average or long national and international calls - and the type of connection and transmission elements, used in the different public and private networks. Therefore, figures 1 to 4 only demonstrate some possible configurations, each one with respect to a specific transmission parameter, which is mainly influenced by this type of network configuration.

Figure 1 illustrates a fully analogue routing between two analogue telephone sets. The most critical parameter in those configurations is the Overall Loudness Rating (OLR). Also noise and frequency response may decrease the overall voice transmission quality.

In figure 2 a digital telephone set is connected with an analogue telephone set and its terminating hybrid. The call is routed via a virtual ATM-connection within the public network. The main impairment in those configurations is caused by a talker echo effect at the digital telephone set, due to the transmission time - increased additionally by ATM - and the signal reflections at the far end hybrid.

Although echo may be suppressed by echo cancelling devices, a further impairment may arise as a result of too high values for the absolute mean one-way delay between the two subscribers. Such a configuration is shown in figure 3 for a call of a mobile telephone (e.g. according to GSM standards) via an international satellite link.

The increasing use of DCME - due to economical reasons - both in public and private networks, as illustrated in figure 4, may contribute high quantizing noise, and with respect to the used coding algorithm, also excessive additional transmission time.

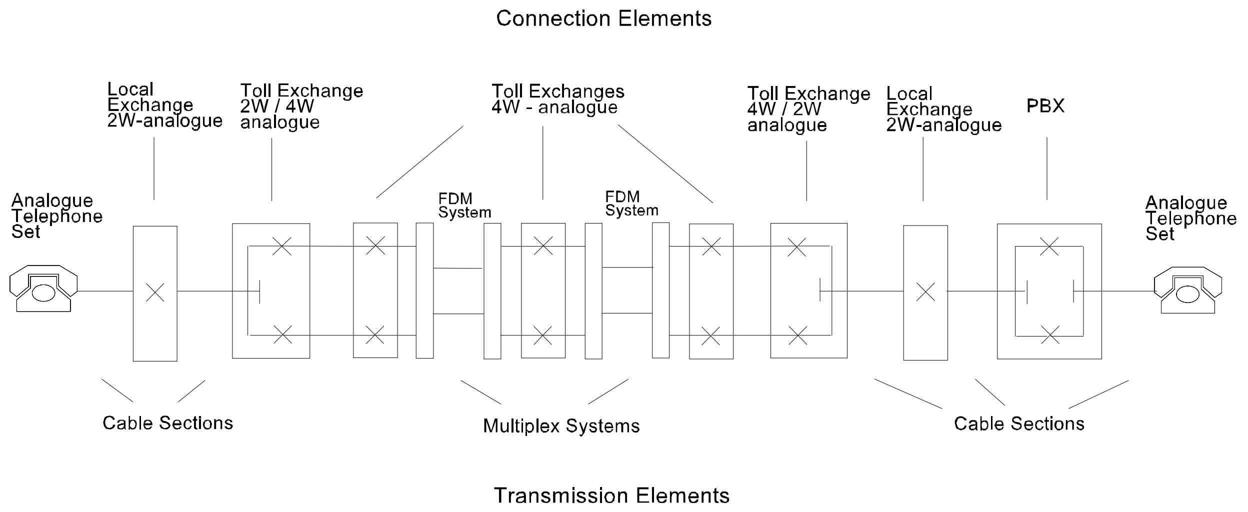


Figure 1: Typical configuration for a fully analogue connection

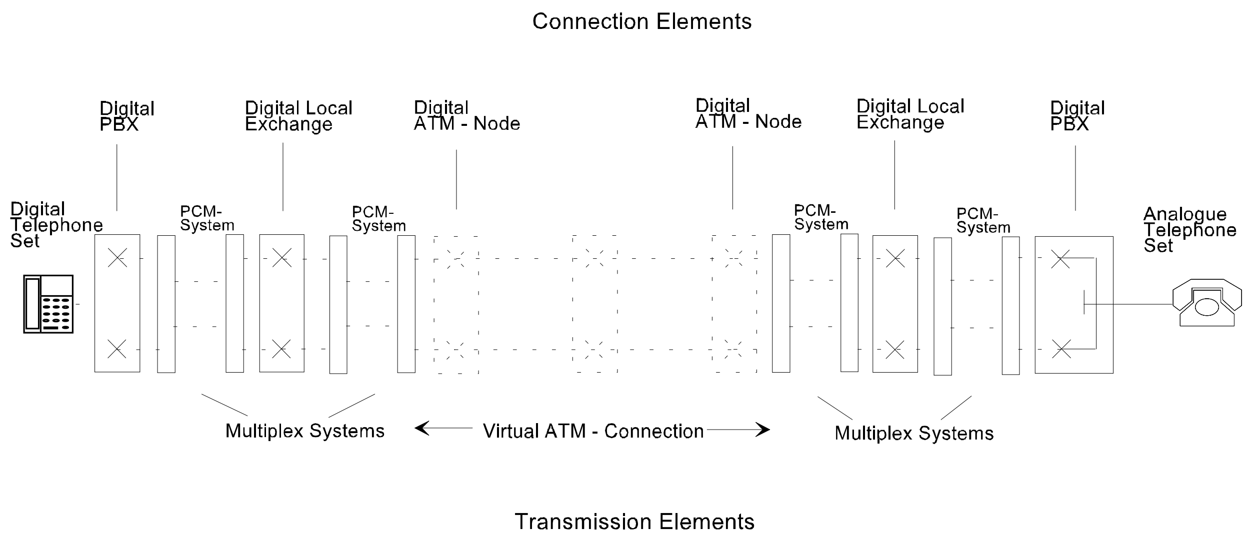


Figure 2: Configuration for a fully digital connection including ATM sections with a terminating hybrid

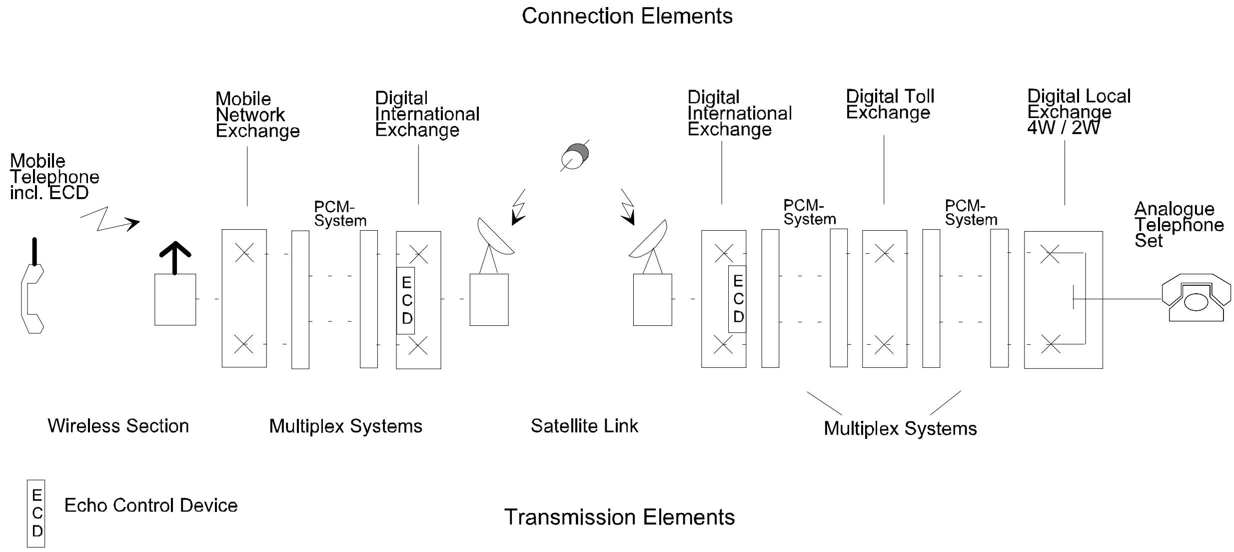


Figure 3: Configuration for a mobile telephone connected via satellite link

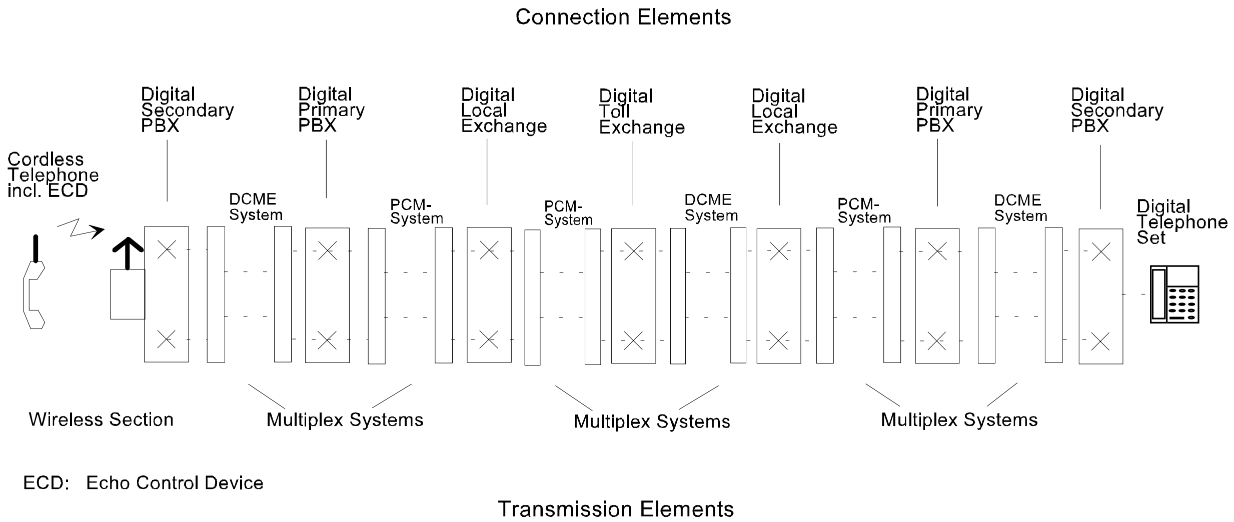


Figure 4: Configuration for a cordless telephone connected via DCME equipment

5 General considerations; individual parameters and combination effects

The ITU-T (formerly CCITT) Recommendations, concerned with speech transmission quality, in general treat each transmission parameter separately, stating individual parameter limits for acceptable performance. This has proved to be a useful approach in many cases. Therefore, this ETR (in clause 6) also lists performance criteria for those individual parameters that may cause transmission impairments.

There are also types of equipment and systems with special transmission properties that need particular consideration with regard to their influence on the overall transmission properties. These are discussed in clause 7.

However, there is also a need to consider the *combination effect* of those impairments that can occur simultaneously in a connection. The "worst case" where all impairments lie at their upper, individually permissible limits would surely result in a very bad transmission quality. On the other hand, there are also cases where one impairment masks (i.e. hides) the effect of another. Furthermore, it is desirable to be able to "trade" impairments so that a connection, where most impairments are quite small, could be allowed one larger impairment and still be considered as good.

The P-series ITU-T Recommendations contain references to some computation models by which combination effects of impairments can be evaluated. Based on this information, a new computation model has been developed and is presented in clause 8. It is believed that this model will help transmission planners to design cost-effective networks with an adequately high transmission quality.

The new computation model includes impairments that were not considered in the other models, like the effect of low bit-rate codecs, short-delay echoes and long absolute delays. A new quality factor, related to users' expectations regarding different types of connections, has also been introduced, see more below.

The structure of the new model is simple, in certain cases simplified, and therefore it is more easy to up-date the model if and when it is found desirable to do so. (For instance when results from new subjective tests and public surveys become available, more experience from application of the model is gained and/or the users gradually change their opinion of what constitutes speech communication quality over different types of connections).

In this context it is worth noticing the modern concept of quality:

- "Quality of a product or a service depends on what the user requires from it.....Quality is simply meeting the customer requirements" (see Total Quality Management [90]);

- "Fitness for purpose or use";

- "The totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs" (see ISO 8402 [147]).

In fact, the user's perception of "a satisfactory voice transmission quality" is not a unique entity but may vary, depending on his expectation of the particular service he is using. One may discern the following system categories:

1) broadcast quality. "True-to-life" reproduction of sounds, Hi-Fi systems;

2) trunk (in US "toll") quality**:** A quality characterised by good intelligibility, good speaker identification, naturalness and with only minor disturbing impairments, for example, the average speech quality of long distance Public Switched Telephone Network (PSTN) connections;

3) communications quality: A quality characterised by good intelligibility, speaker indentity maintained, but with some loss of quality when directly compared with trunk quality, for example, the speech quality present in many mobile communications systems;

4) synthetic quality: Good intelligibility, but the possibility to identify the speaker is mostly lost, and the speech sounds synthetic in nature.

For each category, the user may find an offered system performance quite satisfactory, i.e. having a "good quality" with regard to his needs and expectations. It is only in exceptional cases that it is relevant for the user to make relative quality comparisons between the different categories. (It is worth noticing that a user's perception of what constitutes "good quality" may change with time and experience).

The ETSI computation model is meant to be used for categories "Toll quality" and "Communications quality". To enable comparisons between the two categories with regard to a user's quality perception of the actual service, a concept of an "expectation factor A" has been introduced. To give a clear indication of what is meant, the term "speech communication quality" will be used in this context. However, when the task is to make direct comparisons between systems belonging to respectively category 2 and 3, the expectation factor A is omitted. Then the term "voice transmission quality" can be used.

Note that a user's judgement of a speech communication channel is based on several speech-related perceptual factors such as overall quality, loudness, intelligibility, naturalness. For a telephone type channel, table 1 indicates how these perceptual factors depend on certain physical factors of the channel.

Table 1: Speech perceptual factors and their relation to some telephone channel physical factors

	Perceptual factors			
Physical factors	**Overall quality**	**Loudness**	**Intelligibility**	**Naturalness**
Loss	+++	+++	+++	+
Bandwidth	+++	++	++	+++
Noise	++		++	+
Distortion	++		++	++
Sidetone	+	+		++
Echo (talker)	+++		+	++
Delay	++			+

See also annex J for a discussion of the "Utility" concept as well as some notes on price - quality considerations for telecommunication services. (For obvious reasons, price considerations have not now been included in the model).

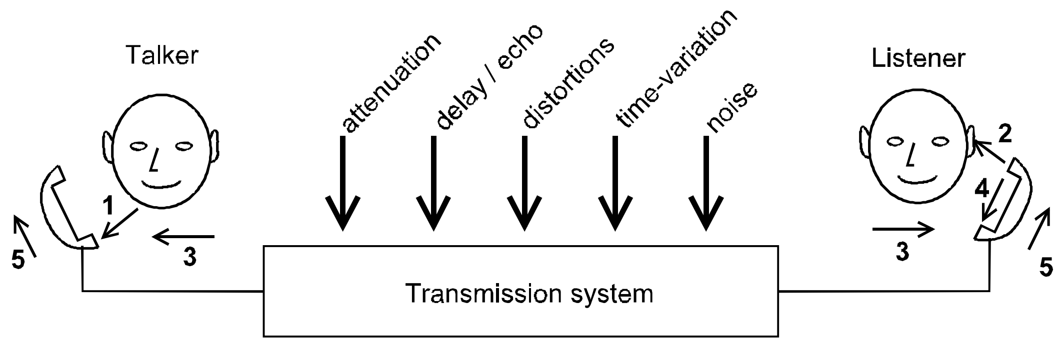
6 Individual transmission parameters

6.1 General survey of the individual transmission parameters

6.1.1 Overview

The aim of this survey is to list individual transmission parameters of importance and give short descriptions of how they influence the total voice transmission quality. The more detailed information is given in subclauses 6.2 to 6.10.

Figure 5 depicts those parameters that influence the overall voice transmission quality from talker to listener. (It is assumed that handsets are used).



1. Sound pressure produced by the talker
2. Transmission path between earphone and eardrum
3. Ambient (room) noise
4. Loss between earphone and microphone
5. Loss between microphone and earphone

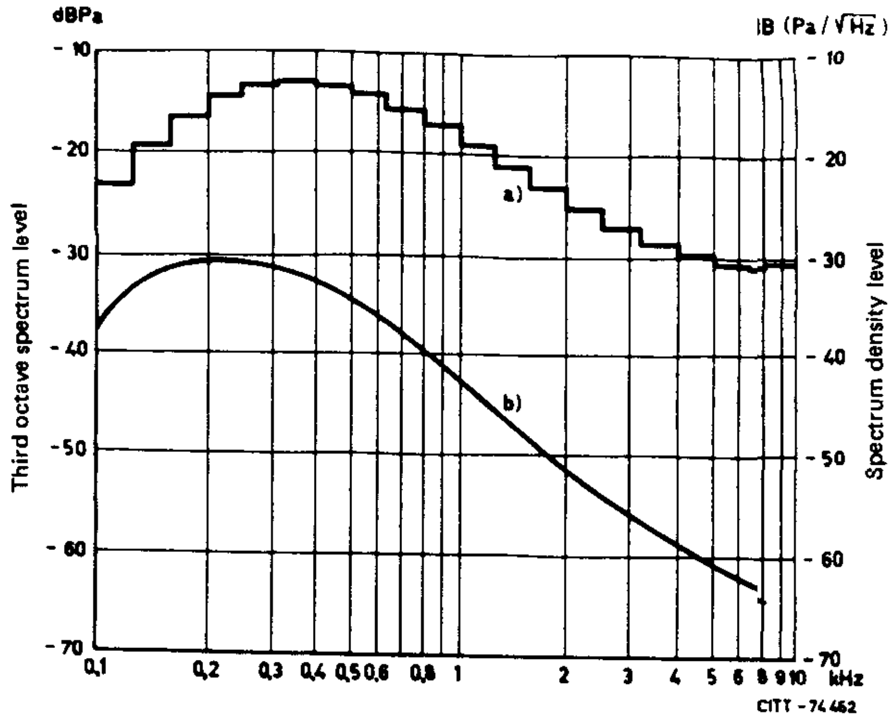
Figure 5: Transmission parameters influencing the quality of a handset telephone connection

6.1.2 The human speech and hearing parameters

The sound pressure levels of speakers vary to a large degree. The level of -30 dBPa has been indicated as the mean sound pressure level of a speaker at a distance of 1 m. At that distance, the levels of 92 % of all speakers lie in the range from -39 dBPa to -19 dBPa. As assumed in the ITU-T P Recommendations, at 25 mm (the "typical" microphone distance) the "normal" speech level is -4,7 dBPa. In practice, the variation between speakers' voice levels and their handset positioning methods give standard variations of 3 dB and 4 dB respectively, resulting in a total standard deviation of about 5 dB around this mean value.

The shape of the voice frequency spectrum also varies with speakers and, to some extent, with the voice level. For measuring purposes, in ITU-T Recommendation P.50 [107] an "artificial voice" is specified, i.e. an electric signal that simulates the long-term spectrum of an "average" human voice. This spectrum is depicted in figure 6.

The instantaneous speech signal is more "peaky" than white noise. Peaks may occur that are 18 dB higher than the active r.m.s. value. However, the speech is not very sensitive to a moderate amount of clipping so that a clipping level 12 dB higher than the r.m.s. value is just noticeable. See subclause 6.8.1 for further details.



a) Third-octave spectrum density level; dBPa
b) Linear frequency spectrum density level; dB(Pa / $\sqrt{\text{Hz}}$)

Figure 6: Average long-term spectrum of the human voice as simulated by the "artificial voice" in ITU-T Recommendation P.50

The human hearing sensitivity for single tones (and narrowband noise) as function of frequency is illustrated by figure 7 (reprinted from ISO 226 [151]). To interpret the curves correctly, the following should be noted:

- the unit "phon" is a measure of the loudness level of tones with the loudness level at 1 kHz defined as the reference. At 1 kHz, the numerical value of phon is the same as the sound pressure level in dBPa (= dB(rel 20 µPa));
- the source of sound is directly in front of the listener;
- the sound field in the absence of the listener consists of a free progressive plane wave and the sound pressure is measured in the absence of the listener;
- the listening is binaural;
- the curves represent "mean hearing sensitivities" of listeners in the age group 18 to 30 years having normal hearing acuity.

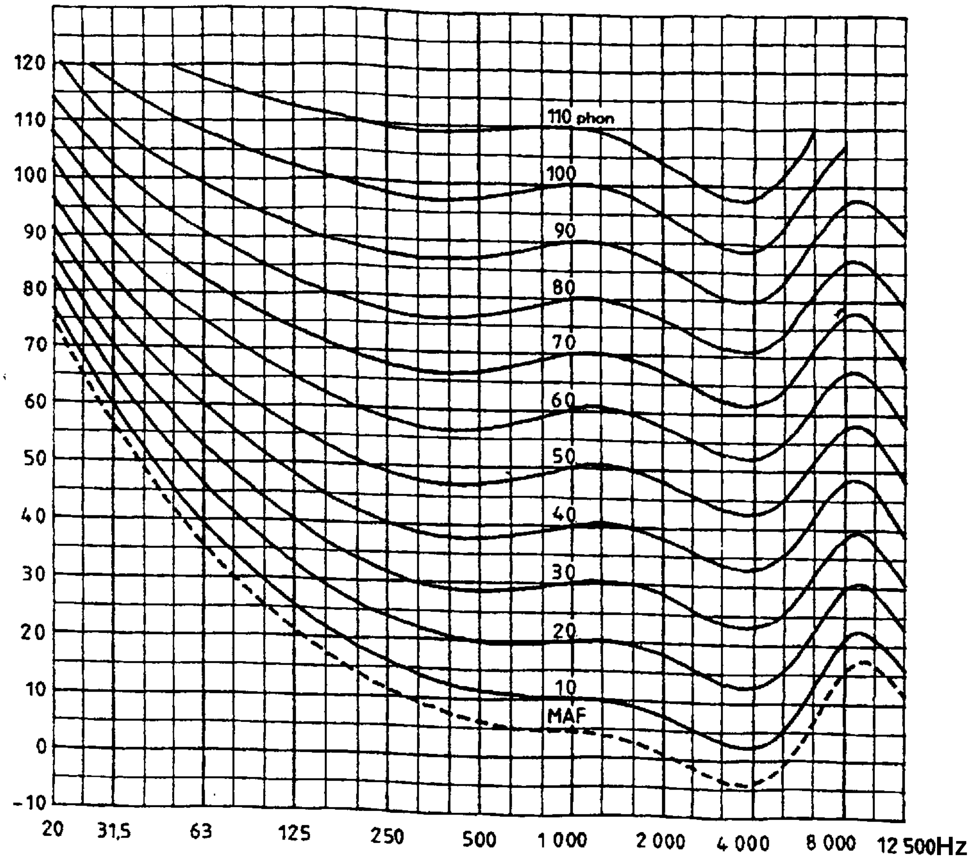
Note that the curves in figure 7 include a certain "sound diffraction effect" caused by the listener's head disturbing the free field of the sound source. However, the difference between the "free field" and the "diffuse field" sensitivities within the telephone speech band are rather minor so that this effect can be ignored.

When listening with only one ear instead of two, the sound impression level at the hearing threshold becomes lower. For a steady state sound and in a free sound field the threshold sound pressure level will be 3 dB higher if the hearing acuity is the same for both ears. If the ears differ slightly in hearing acuity the increase in the threshold will be only about 1,5 dB, see "Telecommunications by Speech" [108].

Going from the threshold levels to normal levels the difference between monaural and binaural listening is greater. For speech the difference in a free sound field is about 10 - 12 dB at normal speech levels. When comparing one-ear handset listening with two-ear loudspeaking listening an additional 2 dB are to be added to take into account the effect of sound diffraction around the head of the listener. The difference between handset telephony and handsfree telephony will consequently be about 14 dB, which is reflected in the ITU-T Recommendation P.34 [109] for handsfree telephony.

The main effect when using two ears is that the brain can detect even small time differences between the sound signals reaching the two ears, even down to 1 ms. This ability allows a person (to a certain extent) to concentrate his listening to a specific direction. This fact lies behind the so-called "cocktail party effect", i.e. the possibility to concentrate on one talker even in a noisy surrounding crowd. It may also explain why a "delayed sidetone" via the handset sounds unfamiliar and peculiar to a talker because the "free ear" is reached directly by the talker's voice (see subclause 6.4.7).

Sound pressure level
dB rel. 20 µPa



NOTE 1: MAF = Minimum audible field; i.e. hearing threshold.

NOTE 2: The frequency scale markings are at 1/3 - octave spacings.

Figure 7: The human hearing sensitivity for single tones and narrowband noise; binaural listening under free field conditions

For *complex sounds,* like a mixture of tones, office room noise, the human voice, etc., the overall loudness impression depends on rather complicated mechanisms, including masking effects between frequency components. However, for speechband telephony applications, when the *sound source is the human voice*, a rather simple model has been found adequate for evaluating loudness impressions in the form of the weighted electro-acoustic losses, "Loudness Ratings", between certain interfaces in the telephone network. (This is more practical than dealing with the absolute voice levels).

In ITU-T Recommendations, the following designations are used for these electro-acoustic losses as referred to the particular interfaces:

- Overall Loudness Rating, OLR, between the speaking subscriber's mouth and the listening subscriber's ear via a connection;
- Send Loudness Rating, SLR, between the speaking subscriber's mouth and an electric interface in the network;
- Receive Loudness Rating, RLR, between an electric interface in the network and the listening subscriber's ear;
- Circuit Loudness Rating, CLR, between two electric interfaces in the network;
- Talker's Sidetone, Sidetone Masking Rating, STMR, between a subscriber's mouth and his (earphone) ear via the electric sidetone path;
- Listener's Sidetone Rating, LSTR, between a room noise source (measured at the position of the handset microphone) and the subscriber's earphone ear via the electric sidetone path;
- Talker Echo Loudness Rating, TELR, the loudness loss of the speaker's voice sound reaching his ear as a delayed echo;
- Listener Echo Loudness Rating, LELR, the difference in loudness loss between the speaker's direct voice sound and its delayed echo reaching the listening subscriber's ear;
- Crosstalk Receive Loudness Rating, XRLR, the loudness loss from a disturbing electric interface to the disturbed subscriber's ear via the crosstalk path.

NOTE 1: In loudness rating contexts, the subscribers are represented from a measuring point of view by an artificial mouth and an artificial ear respectively, both being accurately specified.

NOTE 2: In transmission planning, a convention of loudness rating additivity is applied. i.e. if the circuit between interfaces is subdivided into sections, the sum of the individual section loudness ratings is equal to the total loudness rating.

For further details, see Annex A of ITU-T Recommendation G.111 [34], as well as the P - series of ITU-T Recommendations.

In this model, the ear can be thought of as a bank of bandpass filters approximately equally spaced on a logarithmic frequency scale. If the sound signal in a certain band exceeds the threshold of hearing, the corresponding filter produces an output. All filter outputs are then added to create an impression of loudness, the rule of addition depending on the sound level.

For *very low* sound levels (near the threshold of hearing) the filter outputs are added on a power basis. For *normal* speech sound levels, the loudness measure can be described as obtained neither as power nor voltage addition but rather as the sum of the *logarithm* of the filter outputs. The procedure can be described by equ.(6.1.1) which covers sound levels from very low to normal. The quantity *LR* in this equation is the general form of "Loudness Rating".

$$LR = Lo\frac{10}{m}.\lg\left[\sum_{i=1}^{N} Ki.10^{-0,1mLi}\right] \qquad (6.1.1)$$

where

Lo is a constant, depending on what loudness rating is to be computed.

N is the number of equivalent bandpass filters, the index *i* refers to filter number *i* at frequency *fi*. (Usually, the frequencies *fi* are chosen with a 1/3-octave spacing, starting at 200 Hz and ending at 4 kHz, making *N* = 14).

Li is the loss (in dB) at *fi* of the path under study.

m is the so-called "loudness growth factor", a constant depending on the sound level:

- $m \approx 0{,}2$ for normal speech levels, including sidetones;
- $m \approx 0{,}5$ for "lower" sound levels (corresponding to voltage additions);
- $m \approx 1$ for very low sound levels, near the threshold of hearing, e.g. talker echo (corresponding to power addition).

Ki is the weighting coefficient at *fi*. The *Ki's* have the general properties that their sum is equal to 1 in the frequency range considered:

$$\sum_{i=1}^{N} Ki = 1 \qquad (6.1.2)$$

The *Ki's* are determined by the following factors:

a) the voice spectrum of the "average" speaker;
b) the hearing acuity of the "average" listener;
c) the frequency response of the "nominal" path typical for the particular LR in question.

See also ITU-T Recommendation P.79 [33] and its Annexes.

An interesting fact is that for small values of *m* and a moderate spread between the *Li*-values, equ.(6.1.1) can be simplified to:

$$LR = Lo + \sum_{i=1}^{N} Ki \cdot Li \qquad (6.1.3)$$

This linear approximation is the reason why the total loudness rating of a connection can be computed by simply adding the loudness ratings of its parts. (A rule of thumb: if m is in the order of 0,2 and the spread in *Li* is less than 10 - 15 dB, equ.(6.1.3) can be applied).

NOTE 3: In descriptions of ITU-T loudness ratings, equ.(6.1.1) most often is written in the equivalent form.

$$LR = \frac{10}{m}.\lg\left[\sum_{i=1}^{N} 10^{-0,1m(Wi+Li)}\right] \qquad (6.1.4)$$

As can be seen, *Lo* and *Ki* have been transformed into a term *Wi* in the exponent:

$$Wi = Lo + \frac{m}{10}\cdot\lg\left(\frac{1}{Ki}\right) \qquad (6.1.5)$$

In ITU-T Recommendation P.79 [33] the values of *Wi* are tabulated for SLR, RLR and OLR, using the value *m* = 0,175. For the Sidetone Masking Rating STMR, applicable for characterization of the talker sidetone, *Wi* is tabulated using the value *m* = 0,225.

NOTE 4: The talker echo loudness rating TELR is defined by the convention of additivity, i.e. as

$$TELR = SLR + RLR + Le \qquad (6.1.6)$$

where *Le* is to be interpreted as the echo loudness loss of the electric echo path, and SLR and RLR apply to the interfaces to this path. See subclause 6.5 for further details.

When the sound signal emanates from *a noise source* somewhere in the connection, the evaluation is based on the loudness as such, however with a similar approach as described by equ.(6.1.1). As noise signals are not permitted to reach high levels, the summation is based on power addition. The frequency weighting, which includes the frequency response of a typical handset earphone as well as the sensitivity curve of the ear, is given in ITU-T Recommendation O.41 [32] which describes a suitable weighted noise level measuring instrument, the so-called psophometer.

Note that room noise (ambient noise) is measured by means of special sound level meters conforming to IEC Publication 179 [152] and using the A-weighting. The unit of measurement is dB(A) or, alternatively, dBPa. (X dB(A) = X - 94 dBPa).

For the sake of completeness, it is worth mentioning that active speech levels (r.m.s). are to be measured by a special type of speech voltmeter specified in ITU-T Recommendation P.56 [112].

6.1.3 The voice transmission attenuation between talker and listener

The complete transmission path between the talker and listener consists of several links in tandem as described in subclause 6.2. The input to the chain is described as the "Mouth Reference Point", MRP, and the output as the "Ear Reference Point", ERP, both concepts being defined in ITU-T Recommendation P.64 [40].

The weighted electro-acoustic loss between the MRP and the ERP is characterised by the Overall Loudness Rating (OLR), which is the sum of Send (SLR), Circuit (CLR) and Receive (RLR) Loudness Ratings (see subclause 6.3).

The frequency range for an assured passband of the voice signals is from 300 Hz to 3 400 Hz which is deemed satisfactory for voice telephony. However, in many analogue local networks the lower limit is extended to 200 Hz which considerably improves the naturalness of the received speech.

6.1.4 Talker and listener sidetone

"Sidetone" is the designation for those sounds that are picked up by the handset microphone and transmitted to the earphone of the same handset via the so-called sidetone path which has at most only a minor delay, less than a few ms.

The talker sidetone refers to the pick-up of the talker's own voice and is characterised by the "Sidetone Masking Rating", STMR.

The listener sidetone refers to the pick-up of ambient noise and is characterised by the "Listener Sidetone Rating", LSTR.

See subclause 6.4 for more information.

6.1.5 Echo and stability

When signal reflections occur (e.g. due to impedance mismatches) in association with noticeable delays, talker and listener may be subjected to disturbing echoes.

The talker echo, characterised by the "Talker Echo Loudness Rating" TELR, causes the talker to hear his own voice as an echo. In modern networks, talker echo appears to become the major impairment.

The listener echo, characterised by the "Weighted Echo Path Loss" WEPL, causes the listener to hear an echo of the received speech.

Long echo delays makes the annoyance worse. In addition, if the WEPL value is too low, there is a risk for instability (singing) of the connection.

See subclause 6.5 for further information.

6.1.6 Transmission time

Even in the absence of disturbing echoes, long absolute delays cause communication difficulties for subscribers. See subclause 6.6.

6.1.7 Noise and quantizing distortion

Electric circuit noise used to be the limiting quality factor for long-distance connections but this is hardly now the case with the advent of digital circuits. Note, however, that ambient (room) noise can be picked up both at the send and the receive side and transformed into equivalent circuit noise! Conventional PCM circuits are associated with quantizing distortion, characterised by the "qdu" values. See subclause 6.7 for details.

Impairment caused by the use of low bit-rate codecs cannot very well be predicted by applying qdu figures. Instead, the methodology proposed in the ETSI model (see subclause 9.1.3.5 use of "equipment impairment factor *Ie*), appears more promising.

Moreover, the advent of new switching and communication systems (e.g. ATM, personal and mobile radio systems, DCME, PCME) will introduce in the telephone networks equivalent circuit noise, either stationary or non-stationary -like, due to any kind of environmental noise (traffic, office, vehicle, etc.). This effect, in conjunction with the effects caused by special devices, like very low bit-rate codecs, voice activity detectors, echo cancellers, discontinuous transmission, speech extrapolation, comfort noise insertion, etc., and by random or burst errors in the transmission channel, will be perceived by users as "noise", either impulsive or not. At present, no recommended method exists to measure such "non-stationary noise" and its effect.

6.1.8 Power handling and non-linear distortion

It is important that the voice signal levels in the complete connection are matched to the equipment power handling capabilities in the chain. This is discussed in subclause 6.8.

6.1.9 Crosstalk

Noticeable crosstalk must be avoided in telephone networks. See subclause 6.9 for guidance.

6.2 Basic attenuation/frequency response parameters

6.2.1 General

The electro-acoustic path from the speaking to the listening telephone user consists of several links in tandem to which various degrees of tolerances are assigned with regard to their attenuation/frequency response curves.

The wide-tolerance contributors in the overall loss curve are the send and receive characteristics of the handsets and the attenuation of the analogue subscriber lines. Note that there can be large additional loss variations. Depending on how the speaker holds the handset, the send electro-acoustic loss (i.e. real voice to electric signal loss) varies with a standard deviation of about 4 dB. (The r.m.s. speech voltage has a standard deviation of about 5 dB because the speakers' acoustic level has a standard deviation of about 3 dB). Also, the acoustic leak between the human ear and the telephone receiver shows large variations between individuals, listening conditions and types of handsets.

On the other hand, the losses of the 4-wire parts of the connection, consisting of FDM and PCM systems as well as digital exchanges, are required to lie within close limits.

The reason for this is that while the human ear is not very sensitive to even large variations in the overall frequency response, the 4-wire parts, however, need a close loss control in order to safeguard the network against instability and risk for oscillations.

The overall end-to-end response characteristic is in general considered to be equal to the sum of the responses of the individual links, provided the impedance mismatches are kept reasonably small.

2-wire circuit sections, containing a 4-wire loop, may have different frequency response curves in the two directions which furthermore depend on the terminations. The *nominal* loss curve in *one* direction is obtained by breaking the transmission path in the *other* direction.

In transmission planning, the attenuation/frequency response of a circuit or a circuit section is described by two characteristics:

- firstly, by the so-called "composite loss" at the *reference frequency 1 020 Hz*. This is what is often meant with the term "loss" of a circuit or circuit section. (By composite loss is understood the dB relation between input and output apparent power in mVA);

- secondly, by the voltage loss as a function of frequency, referred to the value at the reference frequency 1 020 Hz.

The nominal loss at the reference frequency between two points within a (ITU-T) circuit can be determined by taking the difference in the (circuit) relative levels at the two points. (The relative levels in this case are determined by reference to the so-called "Transmission Reference Point, TRP").

Note, however, that this method cannot be applied when the points belong to two *different* (ITU-T) circuits. Also worth noticing is that, for practical reasons, a port in an equipment can be assigned *one* relative level when the equipment is considered as a link in a circuit, and a *different* relative level when the equipment is considered on its own with regard to its power handling characteristics. (These matters are discussed in annex C of this ETR and Annex A to ITU-T Recommendation G.100 [114]).

An estimation of the general impedance mis-match effects at a certain interface on the overall attenuation response can be made as follows:

Suppose that the nominal impedance at an analogue interface is *Zo* and that the input impedances of the two circuit sections are *Za* and *Zb* at the interface. The individual responses of the two interfacing circuit sections are measured (or computed) with *Zo* as termination at the interface.

The corresponding return losses against *Zo* are

$$Lra = 20 \cdot \lg\left|\frac{Za + Zo}{Za - Zo}\right| \quad \text{dB} \qquad (6.2.1)$$

$$Lrb = 20 \cdot \lg\left|\frac{Zb + Zo}{Zb - Zo}\right| \quad \text{dB} \qquad (6.2.2)$$

The impedance mismatch causes a difference *Ld* between the actual tandem loss of the two sections on one hand, and the sum of the individually measured losses on the other. This difference lies within the limits

$$20 \cdot \lg\left\{1 - 10^{Lo/20}\right\} < Ld < 20 \cdot \lg\left\{1 + 10^{Lo/20}\right\} \quad \text{dB} \qquad (6.2.3)$$

where

$$Lo = Lra + Lrb \quad \text{dB} \qquad (6.2.4)$$

Table 2 gives some values of *Ld* as function of *Lo.*

EXAMPLE 1: A telephone set is designed for a nominal impedance of 600 Ω resistive with a return loss requirement of *Lra* > 6 dB. The subscriber cable loss is measured (or computed) for a terminating impedance of 600 Ω resistive. In practice, for many lines the input impedance will be capacitive complex and the return loss *Lrb* against 600 Ω resistive is in the order of 6 dB. Thus, *Lo* = 12 dB which results in a difference between the actual tandem response and the sum of the telephone set characteristics and line loss which may lie within the limits

$$-2{,}5 < Ld < 1{,}9 \qquad \text{dB.}$$

EXAMPLE 2: A complex nominal impedance *Zo* is chosen in such a way that it provides a fairly good match to the characteristic impedance of the subscriber cables used. The (modern) telephone set is designed to have the same complex nominal impedance of *Zo* and it may easily achieve a return loss *Lra* > 15 dB against that. Likewise, the return loss of the subscriber cable input impedance against *Zo* will also fulfil *Lrb* > 15 dB. Thus, the corresponding limits are

$$-0{,}3 < Ld < 0{,}3 \quad \text{dB.}$$

EXAMPLE 3: If the two return losses each are 20 dB, the corresponding limits are

$$-0{,}1 < Ld < 0{,}1 \quad \text{dB.}$$

NOTE: When only SLR and RLR are to be measured for a telephone set, it may be permissible to use 600 ohms resistive terminations instead of the correct *Zo*. The reason is that the response errors at low frequencies compensate the errors at high frequencies in the loudness rating calculations.

Equ.(6.2.3) can still be applied if one of the interfacing circuit sections contains a 4-wire loop. The corresponding input impedance is then determined with the 4-wire loop opened. However, the overall attenuation/frequency response will then also have an additional ripple caused by the mismatches between the balance impedances used in the 2-wire/4-wire interfaces and the 2-wire impedances seen at those interfaces. If the Open Loop Loss, OLL, is set to OLL = *Lo*, equ.(6.2.3), giving limits for *Ld,* also applies for the ripple limits. See table 2.

Table 2: Limits for *Ld* as function of *Lo* (See equ.(6.2.3))

Lo dB	*Ld*(min) dB	*Ld*(max) dB
3	-10,7	4,6
6	-6,0	3,5
8	-4,4	2,9
10	-3,3	2,4
12	-2,5	1,9
15	-1,7	1,4
20	-0,9	0,8
30	-0,3	0,3
40	-0,1	0,1

By use of computers it is quite feasible to calculate the exact transmission parameters of a complete connection when the circuit parameters of the constituent parts are known. In general, it is very seldom necessary to do so for the overall transmission from one subscriber to another. For sidetone estimations however, it sometimes may be useful to employ computers for more exact impedance and return loss calculations.

6.2.2 Handset telephone characteristics

The handset send and receive characteristics as function of frequency are measured in accordance with ITU-T Recommendations P.64 [40] and P.65 [28].

In general, many different types of handsets are approved for use in a network. Figures 8 and 9 show for example the average send and receive sensitivity characteristics of 25 sets arbitrarily chosen among those approved for use in Sweden. For comparison, the IRS (= Intermediate Reference System) curves have also been depicted. As can be seen, the curves agree fairly well within the passband 300 Hz to 3 400 Hz. The variations among the sets, however, are quite large, as can be seen in figures 10 and 11.

Note that the receiver characteristics are measured with the handset earcap acoustically sealed to the measuring artificial ear. In actual handset use there exists an acoustic leak between the earcap and the human ear. This type of leak can introduce a very large and variable attenuation/frequency distortion, examples of which are shown in figure 12. (This figure refers to a conventional receiver. Receivers with a *high* acoustic impedance will cause even more severe low-frequency cut-offs. Receivers with a *low* acoustic impedance have much better low-frequency responses).

How to include measuring methods for handset acoustic leaks in the ITU-T Recommendations are under study.

The wide variations found in the actual low frequency receiver and leak response curves may cause a corresponding variation in how certain psophometric values of power line interference on a subscriber line is perceived subjectively. (The psophometer weighting includes the receiver sensitivity of a "typical" handset under sealed conditions).

For handsets used in conjunction with some types of *low bit-rate codecs,* such as for mobile applications, a flat frequency sending response may give a slightly better speech transmission quality than that obtained with an IRS characteristic.

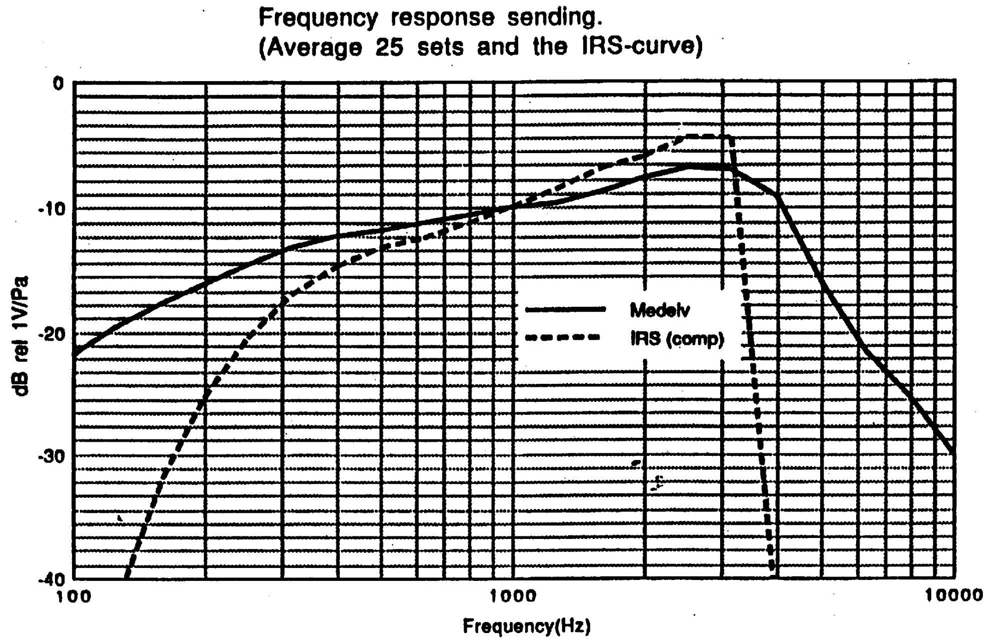


Figure 8: Frequency response sending, IRS and mean value of 25 approved Swedish handsets

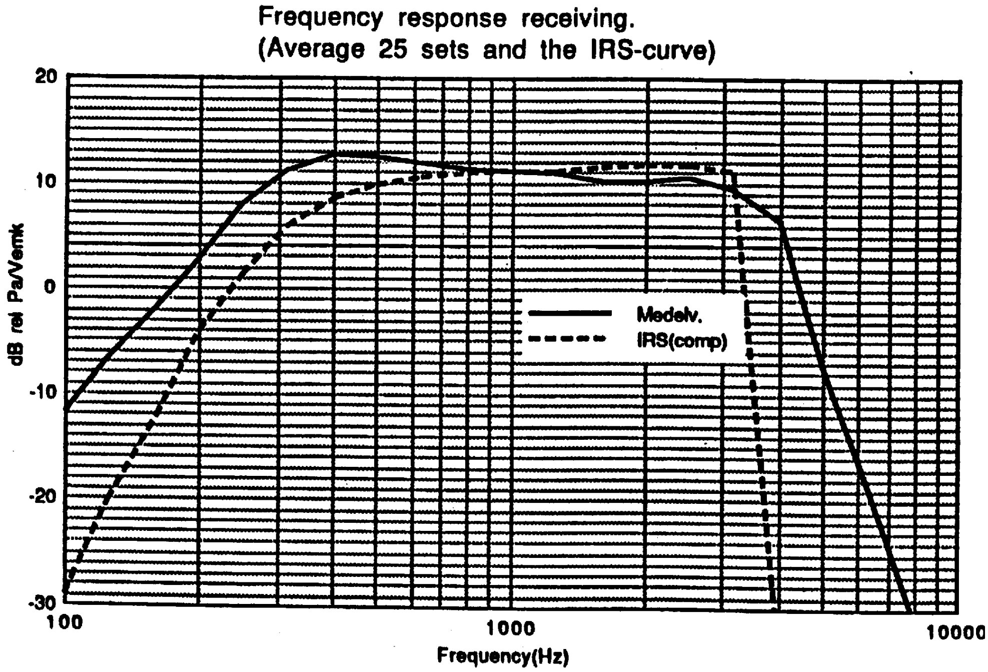


Figure 9: Frequency response receiving, IRS and mean of 25 approved Swedish sets

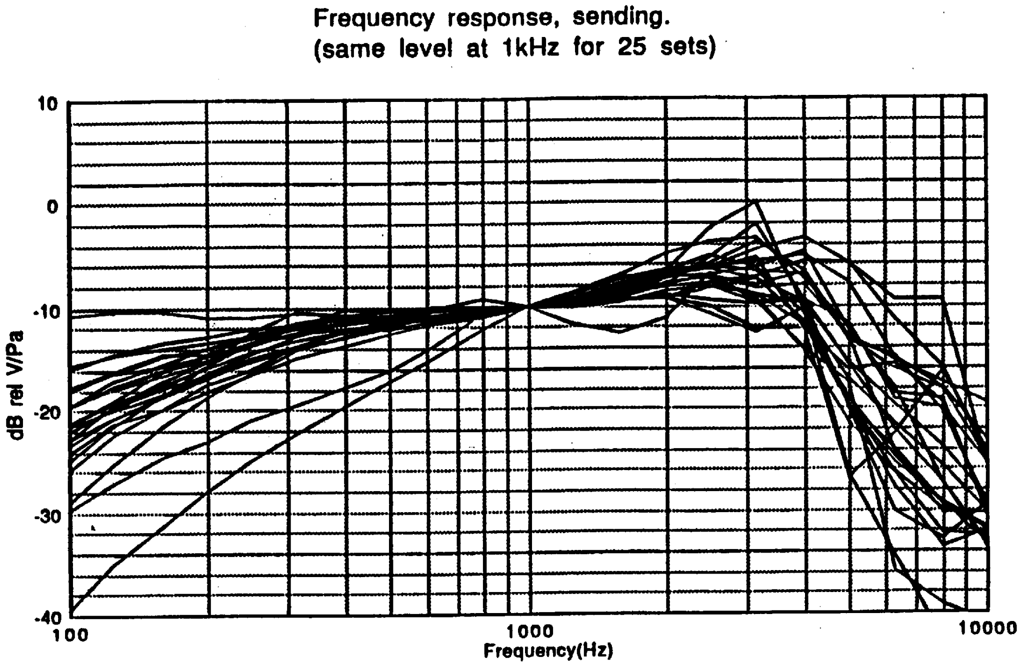


Figure 10: Frequency response, sending. The level is adjusted to
-10 dB relative to 1 V/Pa at 1 000 Hz for all 25 set

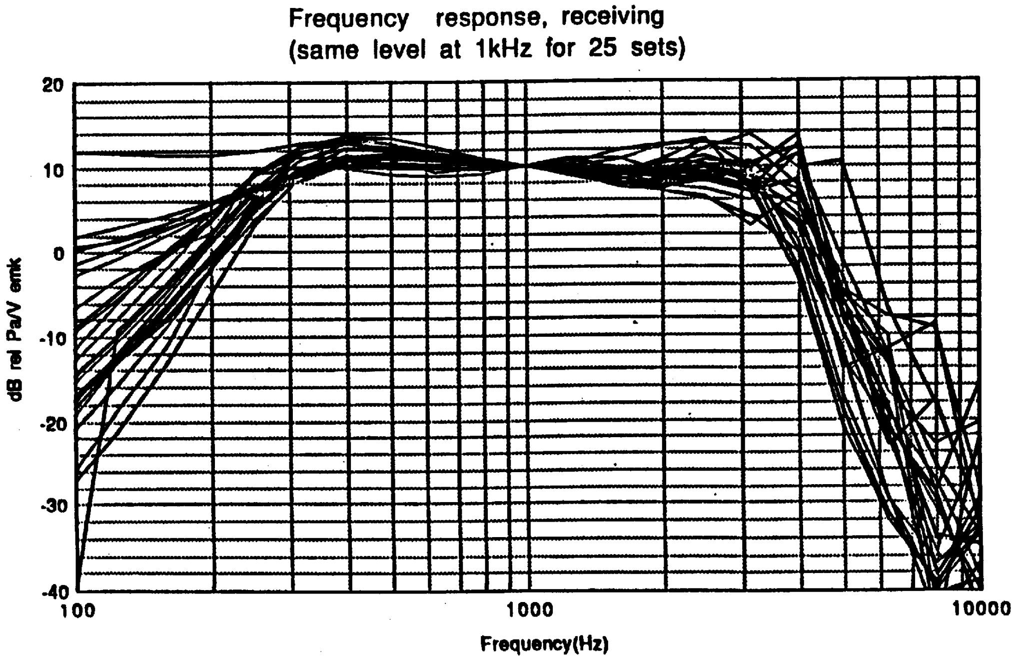


Figure 11: Frequency response, receiving. The level is adjusted to -10 dB relative to 1 Pa/V e.m.f. at 1 000 Hz for all 25 sets

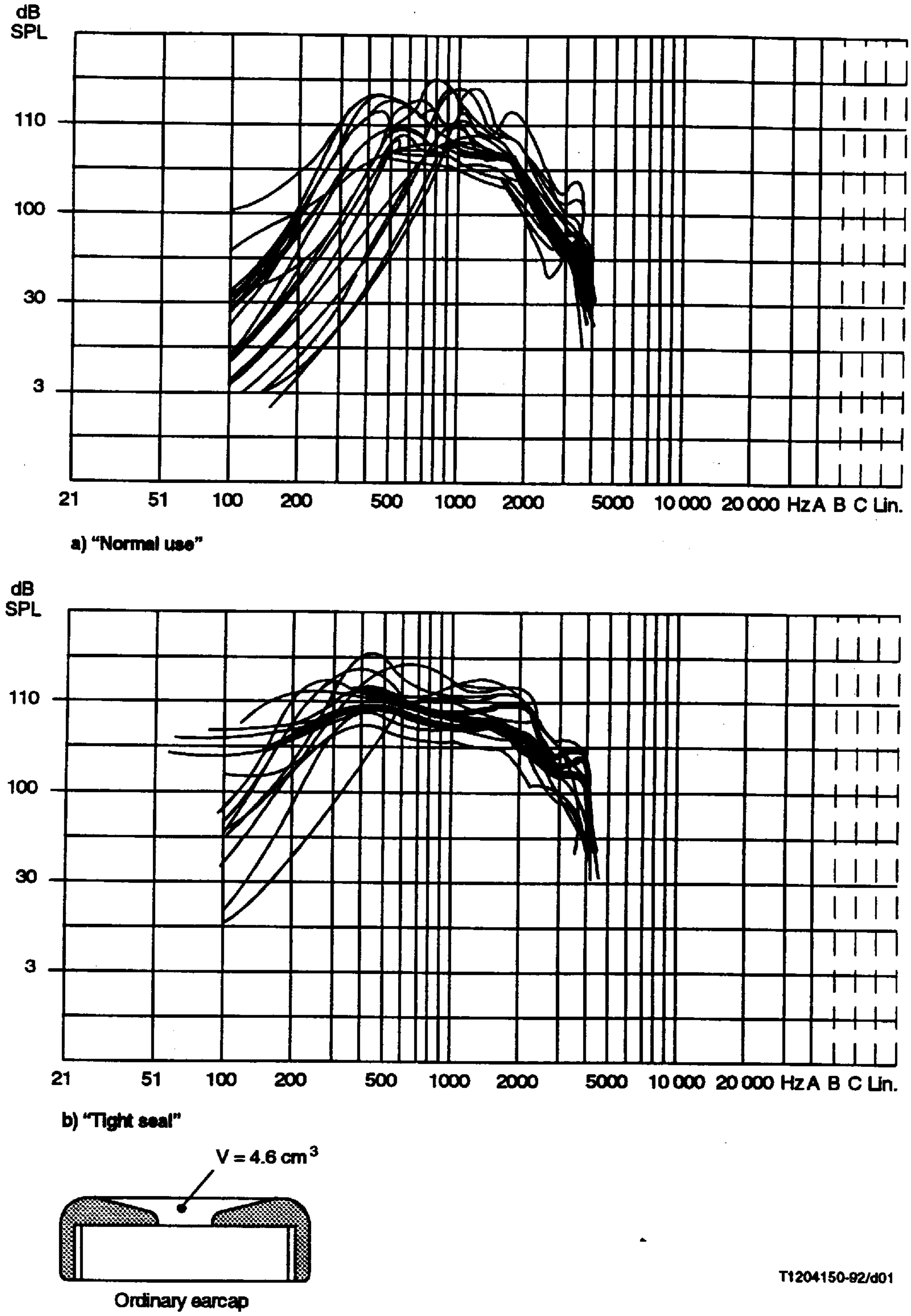


Figure 12: Examples of acoustic leaks between a handset earcap and the human ear (20 persons)

6.2.3 Subscriber line attenuation

The analogue subscriber lines consist almost exclusively of unloaded cables. Their losses vary with the cable parameters, the terminations and the lengths. Methods for calculation of the loss/frequency characteristics are well known and easy to implement in computer programs. Figure 13 shows as an example the loss curve for a typical cable, terminated by a complex impedance.

If the telephone sets have an automatic sensitivity regulation, the loss at the reference frequency is more or less compensated. The attenuation/frequency distortion remains, however. (It may amount to 10 dB for long lines).

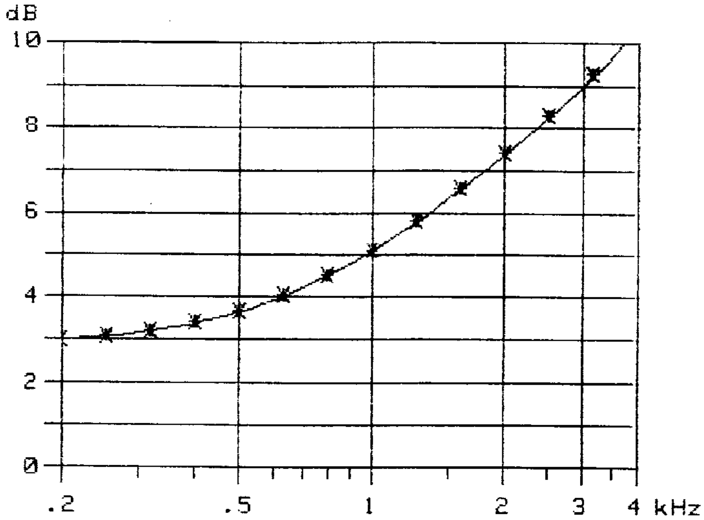


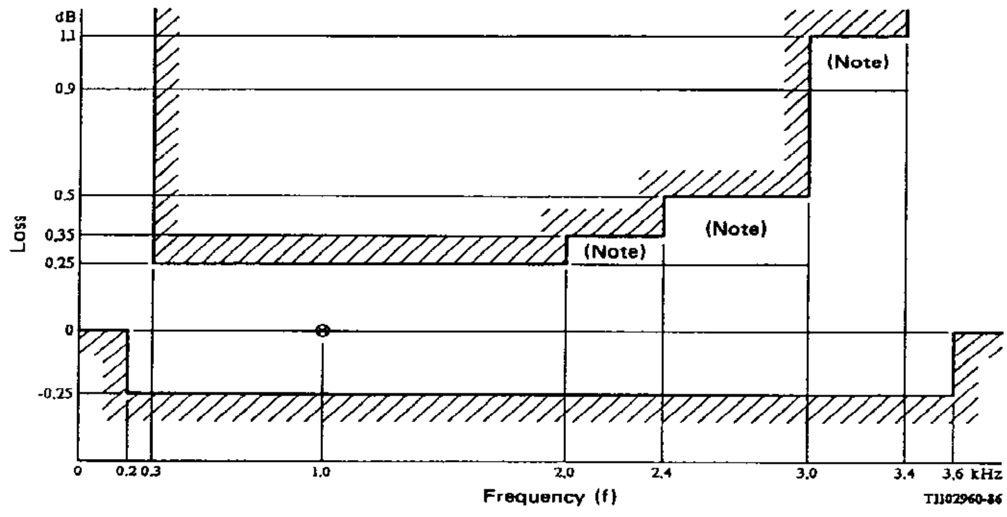
Figure 13: Typical attenuation of an unloaded cable (3 km; 0,4 mm copper; 50 nF/km) Terminations: 270 Ω in series with a parallel combination of 750 Ω and 150 nF

6.2.4 4-wire circuits

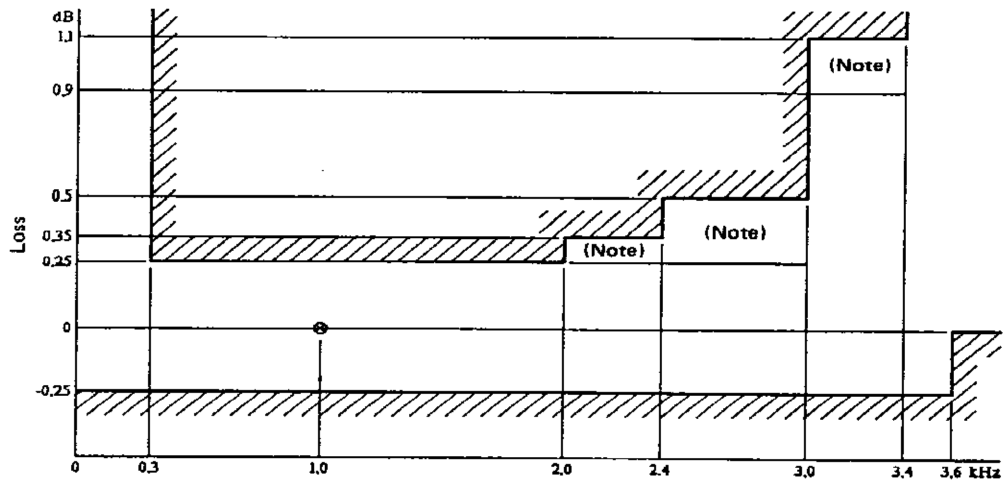
4-wire circuits could in the past be formed by a substantial number of FDM (analogue) systems in tandem. For such configurations the passband loss variations added directly. In order to safeguard against the open loop loss becoming too low or even negative at some frequency with resulting near-singing or instability conditions, it was found necessary to introduce 0,5 dB loss per circuit and allow only minor variations in the nominally flat passband loss curves. The same principles were followed when a number of PCM systems as "digital islands" are interconnected with analogue systems. It turned out, fortunately, to be fairly easy to achieve the tight passband tolerances.

Now, in modern networks, more and more connections with all-digital circuits are used so that virtually only one encoding-decoding process occurs in a connection. However, the close control of the passband response has been kept. Figure 14 shows as an example the half-channel requirements of a digital exchange according to ITU-T Recommendation Q.553 [31].

The influence of the 4-wire circuits on the overall mouth to ear response is in practice the same as an ideal filter with the passband 300 Hz to 3 400 Hz. The passband loss variations can be ignored for voice transmission, provided the open loop losses are high enough.



a) Input connection



b) Output connection

NOTE: In the marked frequency ranges relaxed limits are shown which apply if the maximum length of in-station cabling is used, i.e. 100 m between the distribution frame (DF) and the exchange.

Figure 14: Half-channel requirements of a digital exchange (ITU-T Recommendation Q.553 [31])

The Open Loop Loss *OLL* is equal to the sum of all losses and gains within an (opened) 4-wire loop. The value of *OLL* determines the amount of attenuation ripple *Lripple* in the pass band which will occur when the 4-wire loop is closed. It turns out that

$$Lripple = -20 \cdot \lg\left|1 - K\right| \quad \text{dB} \tag{6.2.5}$$

where

$$K = 10^{-OLL/20} \cdot \left\{\cos B - j \sin B\right\} \quad \text{dB} \tag{6.2.6}$$

and *B* is the phase shift in the open loop. (Note that *OLL* and *B* are functions of frequency).

For high values of *OLL*:

$$\left|Lripple\right| = \frac{20}{\ln 10} \cdot 10^{-OLL/20} = 8,7 \cdot 10^{-OLL/20} \quad \text{dB} \qquad (6.2.7)$$

The effect of the passband ripple on a voiceband signal passing the circuit can be interpreted as follows:

When the 4-wire loop is closed, the signal is subjected to a (complex) attenuation factor

$$\frac{1}{1-K} = 1 - \frac{K}{1-K} \qquad (6.2.8)$$

In the right-hand side of equ.(6.2.8), the first term represents the direct signal as it would pass the circuit with the 4-wire loop opened in the back direction. The second term represents a double-reflected signal, the so-called listener echo, occurring when the 4-wire loop is closed. Compared to the direct signal, the double-reflected signal is attenuated with

$$LE = 20 \cdot \lg\left|\frac{1-K}{K}\right| \quad \text{dB} \qquad (6.2.9)$$

For the high values of *OLL* which often occur in practice:

$$LE \approx 20 \cdot \lg\left|\frac{1}{K}\right| = OLL \qquad (6.2.10)$$

See also subclause 6.5.2 for a discussion of listener echo.

6.3 Overall, send, receive and circuit loudness ratings

6.3.1 General

As used in transmission planning, loudness rating is an objective measure of the loudness loss, i.e. a weighted, electro-acoustic loss between certain interfaces in the telephone network. For more information about definitions and explanations, see clause 3.

Figure 15 shows the loudness ratings (LRs) in a normal speech connection. The designations are:

OLR = Overall Loudness Rating;
SLR = Send Loudness Rating;
RLR = Receive Loudness Rating;
CLR = Circuit Loudness Rating.

In figure 15, CLRn refers to the Circuit Loudness Rating of circuit number n. Furthermore, SLR(set) and RLR(set) apply to the telephone sets themselves. At interface number *No* we have

$$SLR = SLR\,(set) + \sum_{n=1}^{No} CLR_n$$

$$RLR = RLR\,(set) + \sum_{n=No+1}^{N} CLR_n$$

$$OLR = SLR + RLR$$

The Circuit Loudness Rating CLR is equal to the l*oss at the reference frequency 1 020 Hz.* This is of course always true if the circuit has a flat frequency response but it is also valid (with good accuracy) for unloaded subscriber cables.

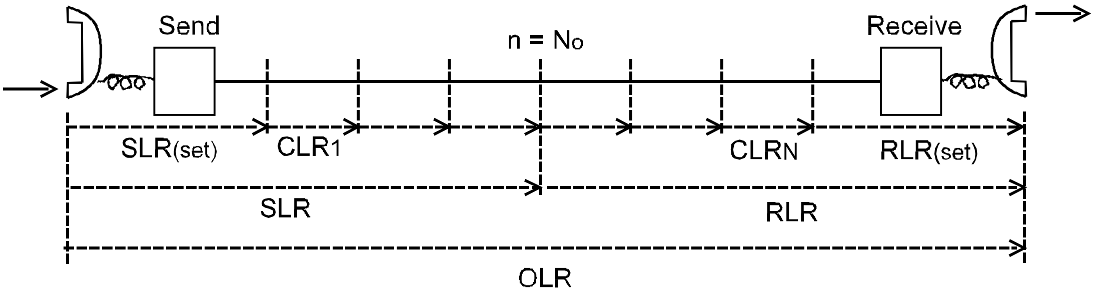


Figure 15: Loudness Ratings in a normal speech connection

The CLR of an unloaded subscriber cable can also be estimated by the relation

$$CLR = K\sqrt{R \cdot C} \qquad \text{dB/km}$$

where

R = Cable resistance in ohm/km;
C = Cable capacitance in nF/km;
K = 0,014 when terminations are 900 Ω resistive;
K = 0,015 when terminations are 600 Ω resistive;
K = 0,016 when terminations are capacitive complex impedances.

6.3.2 Overall Loudness Rating (OLR)

Recommendation G.111 [34] states the following objectives (for traffic-weighted mean values):

Long - term: 8 < OLR < 12 dB
Short - term: 8 < OLR < 21 dB

NOTE 1: The ranges stated for OLR are for planning and do not include measuring and manufacturing tolerances.

NOTE 2: The difference in nominal values of OLR between the two directions of transmission should not exceed 8 dB, preferably not 6 dB.

6.3.3 Send and Receive Loudness Ratings (SLR and RLR)

Recommendation G.121 [38] states the following objectives (for traffic-weighted mean values) which are applicable for the 0 dBr "Virtual International Connecting Point" (VICP). However, they can also be considered as valid (within reason) for other 0 dBr points, such as at a digital bit-stream, in a connection.

Long - term:	7 < SLR < 9	1 < RLR < 3 dB
Short - term:	7 < SLR < 15	1 < RLR < 6 dB
Maximum values for an average-sized country:	SLR < 16,5 dB, RLR < 13 dB	
Minimum value:	SLR > 2 dB	

NOTE: The maximum values of SLR and RLR should preferably not be allowed to occur on the same connection.

6.3.4 Conversions to and from other loudness loss related values

Conversions between ITU-T loudness ratings and others, such as Corrected Reference Equivalence (CRE), OREM, IEEE, etc., should preferably be done by actual measurements on the particular telephone set concerned. However, the following approximate relations are given as a guidance:

CRE: SCRE = SLR + 5; RCRE = RLR + 5;

IEEE: TOLR = SLR - 56; ROLR = RLR + 50; OOLR = OLR - 6;

6.4 Sidetone

6.4.1 General

When a subscriber is engaged in a telephone conversation via a handset, his handset listening ear is not completely shielded but can be reached by sounds emanating from his own surroundings, i.e. both his own voice and the ambient noise. The transmission paths to the handset ear are somewhat different for the voice and the ambient noise, but both signals are classified as sidetone signals. The "own voice transmission" is called "Talker Sidetone" and the "ambient noise transmission" is called "Listener Sidetone". (Note that in a telephone handset conversation, sounds reaching the free ear are more or less "switched off" by the hearing mechanism of the brain).

The talker sidetone path should ideally emulate that transmission path to the talker's own ear which exists in a face-to-face conversation situation. This implies that the talker sidetone loss must lie within certain limits for a comfortable talking situation. On the other hand, the listener sidetone loss must be higher than a certain minimum value to avoid disturbance from ambient noise.

A sidetone path consists in principle of three transmission links in parallel, namely an electric path, an acoustic airpath and, via the handset, a mechanical path.

The electric path has its input at the microphone, passes through the telephone via a special sidetone coupling, which can be external or internal, between the transmit and receive side, and has its output at the earphone. In a conventional, two-wire analogue set, the sidetone coupling occurs in a hybrid circuit as a slight mismatch between the subscriber line impedance seen at the terminals of the set, and the so-called sidetone balance impedance of the set. In a four-wire, digital set the sidetone coupling is achieved by an internal circuitry.

For the talker sidetone, the acoustic path is mainly via bone conduction in the talker's own head. (Additionally, some sidetone sounds may be carried by mechanical transmission - resonances - in the handset to the subject's ear but this is restricted to some very narrow frequency bands).

For the listener sidetone, the acoustic path consists of the acoustic leak between the handset earphone and the talker's ear.

Talker and listener sidetone are characterised in the form of Loudness Ratings, the former as Sidetone Masking Rating (STMR) and the latter as Listener Sidetone Rating (LSTR).

In the subclauses that follow more details will be given of the sidetone subject.

6.4.2 Electric and acoustic paths of the sidetone, calculation of STMR and LSTR

The loss of an electric sidetone path is a function of frequency. For an *analogue* set it can be written as

$$Ast = Lbr - St - Sr \qquad (6.4.1)$$

where

$$Lbr = 20 \cdot \lg \left| \frac{Zo + Zb}{2Zo} \cdot \frac{Z + Zo}{Z - Zb} \right| \quad \text{dB} \qquad (6.4.2)$$

Zb is the sidetone balance impedance, *Zo* is the input impedance of the set, *Z* is the impedance seen at the telephone set terminals.

The acoustic and mechanical sidetone couplings will be noticeable in *Zb* and *Zo* as small changes from their purely electrical values so that equ.(6.4.2) is exact. In practice, however, only the electrical values of *Zb* and *Zo* are used.

St is the transmit sensitivity of the set and *Sr* the receive sensitivity of the set as measured according to ITU-T Recommendation P.64 [40], however with matching terminations as described in annex B of this Recommendation.

For a digital set, a relation similar to equ.(6.4.1) exists. However, the corresponding sidetone loss *Lbr* is set by design and is (in general) made constant with frequency. Moreover, the "send" and "receive" circuitry for sidetone are often different from those for the normal voice transmission.

Note that for talker sidetone the transmit sensitivity *St = St(dir)* applies for direct sound while for the listener sidetone *St = St(diff)* relates to a diffuse sound field; the difference between the two is (most often) termed *DELSM.*

$$DELSM = St(diff) - St(dir) \tag{6.4.3}$$

In general, *DELSM* is rather constant with frequency.

By comparison between typical electric and acoustic sidetone path losses it has been found a significant difference in frequency response for the two. The attenuation of the acoustic sidetone path, whether via bone conduction or earcap leakage, is quite low for low frequencies but is appreciable for higher frequencies. Thus, the acoustic sidetone dominates over the electric sidetone at low frequencies, i.e. below about 600 Hz - 800 Hz. The opposite is true for higher frequencies. Interestingly, the bone conduction path for talker sidetone has about the same frequency response as the typical earcap leakage for the listener sidetone. These facts have some consequences for how STMR and LSTR are to be computed. (The name STMR = Side Tone Masking Rating was in fact chosen in recognition of the masking effect by the acoustic sidetone on the electric sidetone. Note that in the Side Tone Reference Equivalent STRE, the previously used measure for talker sidetone, this masking effect was not considered. Therefore, one cannot state a general relation between STRE and STMR).

The general algorithm for calculation of a Loudness Rating, LR, is of the form

$$LR = \frac{10}{m}.\lg\left\{\sum_{i=N1}^{N2} 10^{-0,1m(Ai+Wi)}\right\} \tag{6.4.4}$$

where for STMR and LSTR:

- m = 0,225;

- the summation is to be performed at frequencies *Fi*, spaced 1/3 octave apart, starting at 100 Hz and ending at 8 000 Hz;

- *Wi* is a weighting coefficient, given in table 3 of ITU-T Recommendation P.79 [33]. *Wi* includes an estimated influence of the acoustic leakage path;

- *Ai* is the loss of the sidetone path under consideration at frequency *Fi.* The acoustic path is not considered. (In a measuring set-up, one determines the sensitivity *Si* = -*Ai*). For computation of STMR and LSTR for analogue sets, *Ai = Ast* and equations (6.4.1) and (6.4.2) apply, and one usually ignores the effect of the mechanical path.

As can be seen in table 3 of ITU-T Recommendation P.79 [33], for low frequencies the weighting coefficients *Wi* are much larger than those above 800 Hz. Thus, applied in equ.(6.4.4), this shows that the electric sidetone loss in the lower frequency band has very little influence on the value of STMR or LSTR due to the dominance of the acoustic sidetone path.

Usually, STMR and LSTR are determined by direct measurement of the sidetone path sensitivity. For analogue sets, this must be done with a representative set of terminating impedances *Z*. Note that even a direct measurement represents an approximation to the actual sidetone conditions which are found in practice. Firstly, because the real earcap leakage is replaced by an assumed one. Secondly, for analogue sets, that the actual terminating impedances are only approximately represented by the terminating impedance in the measuring set-up.

For transmission planning purposes, STMR can also be calculated from the SLR and RLR of the set and a weighted average of *Lbr*. See annex A, equ.(A.5).

6.4.3 Relation between STMR and LSTR

When a telephone set is connected into the telecommunications network there is a firm relation between STMR and LSTR

$$D = LSTR - STMR \tag{6.4.5}$$

The *D*-factor depends only upon the design of the handset, i.e. the shape of the handset and the microphone circuitry. For linear circuits, *D* is independent of the speech level. For non-linear microphone circuits *D* may vary with the speech level, thereby introducing a suppression of the pick-up of ambient noise. In principle, *D* is a weighted average of *DELSM*.

For a *typical, linear microphone* the following values are given as a rough guide:

- conventionally shaped handset of normal length D = 3
- short handset D = 0
- very short handset D = -3

As a matter of fact, for handsets of "conventional form" and with linear microphone circuitry, the factor *D* can be estimated by equ.(6.4.6), using the distance *d* mm between the speakers mouth and the microphone. However, for oddly shaped handsets the deviation from the results of the formula can be relatively large.

$$D = 33 - 20 \cdot \lg(d) = 20 \cdot \lg\left(\frac{45}{d}\right) \tag{6.4.6}$$

(More accurately, the distance *d* is determined in the SLR measuring set-up as the distance between the centre of the external opening for the microphone on the surface of the handset and the centre of the lip-ring of the artificial mouth).

To increase the *D*-factor, *non-linear microphone circuits* may be used, such as carbon microphones or special electronic circuits. Then the *D*-factor is dependent on the room noise level, a threshold effect being noticeable. Thus, for a typical carbon microphone handset *D* is in the order of 6 dB to 8 dB for a (Hoth-type) room noise of 60 dB(A). (For other noise levels and some handset designs *D* can be as high as 15 dB).

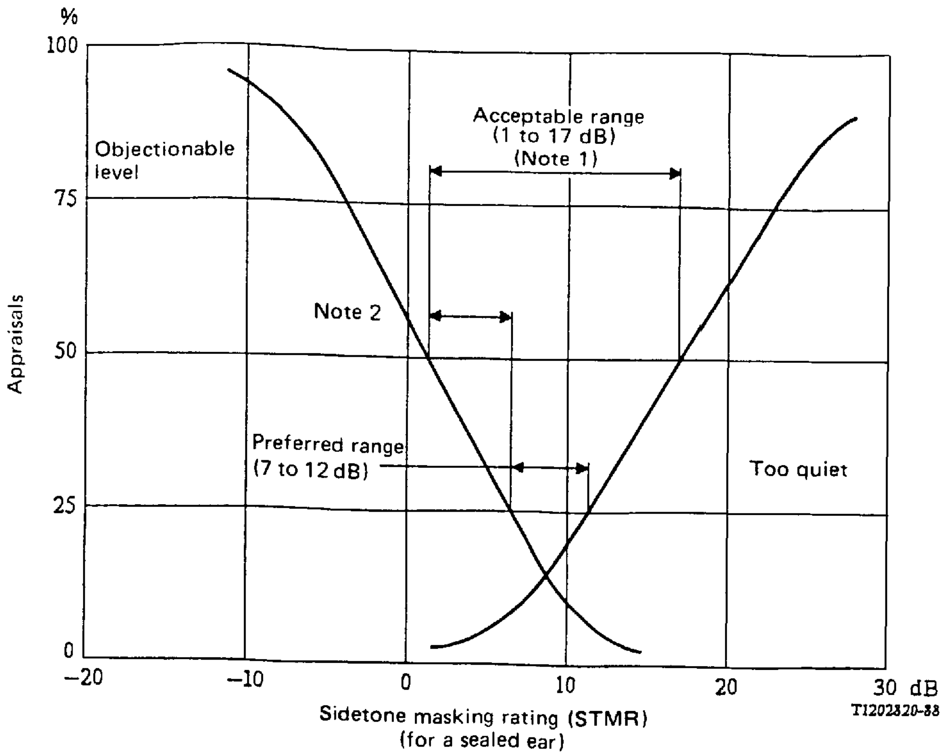
In general, subscribers prefer sets with linear microphones because the sound quality is much superior. However, when replacing old carbon microphone sets in noisy environments with modern linear sets, care must be taken to ensure that the LSTR value is sufficiently high (see subclause 6.4.5).

Also, handsets with linear *directional microphones* may achieve high values of *D*, although with possible difficulty for the talker to position the handset correctly.

The *D*-factor and its measurement are discussed in ITU-T Recommendation P.64 [40].

6.4.4 Recommendations for STMR

The range for acceptable values of STMR is fairly wide. Figure 16, reprinted from Supplement 11 of ITU-T P-series of Recommendations [41], illustrates this. Unfortunately, in many present 2-wire connections the impedance deviations from the ideal are so large that the electric sidetone feedback becomes too strong, i.e. STMR too low. This causes the speaker to lower his voice and/or move the earphone away from his ear, thus impairing the acoustic transmission quality.



NOTE 1: Conversational conditions will determine what part of this range is acceptable for a given connection.

NOTE 2: This part of the acceptable range (1 dB to 7 dB) should only be entered with caution, e.g. on low loss connections or where there is a receive volume control.

Figure 16: Opinion ratings of STMR, the talker sidetone

ITU-T Recommendation G.121 [38] gives the following values as a guide for transmission planning.

For 2-wire (analogue) telephone sets:

- STMR = 7 - 12 dB Preferred range;
- STMR = 20 dB Upper limit, above which the connection feels dead;
- STMR = 3 dB Lower limit, acceptable only for low-loss connections, i.e. low OLR;
- STMR = 1 dB Lowest short-term limit for exceptional cases (very short subscriber lines).

ITU-T Recommendation P.31 [39] and the revised ITU-T Supplement No. 11 [41] recommend for digital (4-wire) telephone sets:

- nominal STMR = 10 - 15 dB

NOTE 1: When STMR = 7 or 8 dB, this corresponds to the average acoustic loss from the talker's mouth to his ear via the electric sidetone path being about 0 dB in typical cases.

NOTE 2: STMR has to be determined for the *complete* connection.

NOTE 3: In the presence of high room noise, requirements on LSTR may be the controlling factor.

NOTE 4: If the reflected signal has a noticeable delay it is interpreted as an echo rather than talker sidetone, which means it needs more suppression to avoid subscriber dissatisfaction (see subclause 6.4.7).

A loud talker sidetone (i.e. a very low value of STMR) causes the talker to lower his voice so that the listener at the other end may interprete this as if the connection has too much loss. This is illustrated by figure 17 which is reprinted from Supplement 11 to the ITU-T P Recommendations [41].

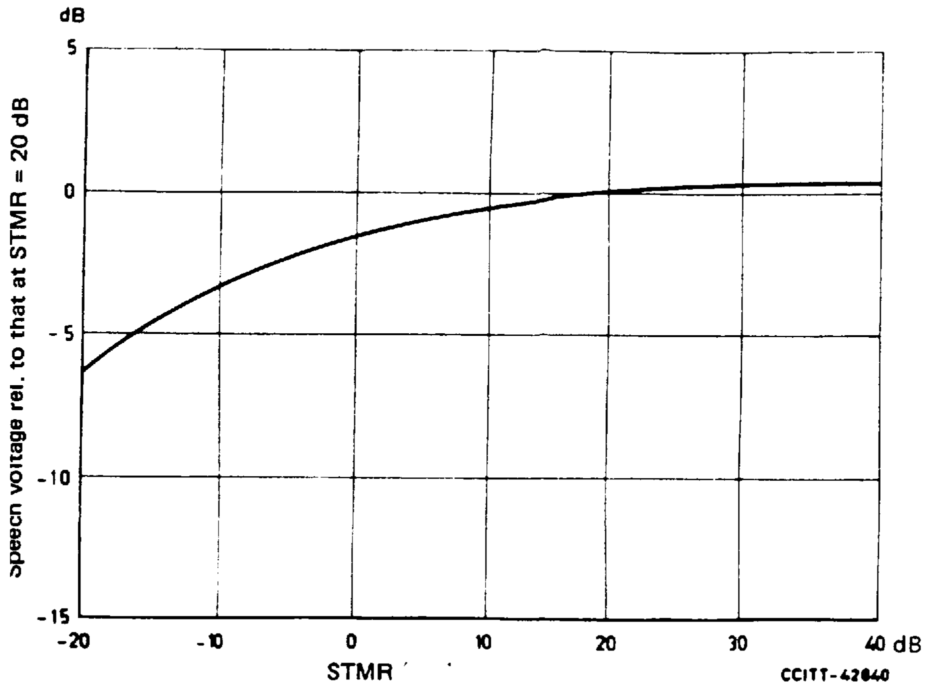


Figure 17: Speech voltage as function of STMR (Connection with OLR near its optimum value)

6.4.5 Recommendations for LSTR

ITU-T Recommendation G.121 [38] states that in modern telephone systems one should strive for

- LSTR > 13 dB.

ITU-T Recommendation P.31 [39] and Supplement No. 11 to the ITU-T P Recommendations [41] recommend for digital sets:

- nominal LSTR > 15 dB.

At this value of LSTR, the earcap provides a certain shielding effect for the ambient noise (see also subclause 6.7.4).

Sometimes the network conditions are such that at some places sensitive analogue telephone sets (i.e. with low SLR and RLR) have to be used. Then in order to fulfill the requirement LSTR > 13 dB it is necessary to provide a quite good matching between the equivalent sidetone balance impedance *Zb* of the sets and the line impedance *Z* (see annex A.7 for a discussion).

6.4.6 Sidetone distortion

Overloading in the sidetone path and the resulting excessive sidetone distortion may cause some very unpleasant sound effects. Therefore, in TBR 8 [45] it is stated that the sidetone distortion should be less than 10 % for an acoustic input of -4,7 dBPa.

6.4.7 A comparison between talker sidetone and talker echo

If the sidetone is delayed longer than about 2 ms mean one-way (i.e. 4 ms round-trip), it begins to be interpreted as a talker echo, even though it may not be perceived as a distinct echo. Low STMR values then result in noticeable speech quality impairments. As an example see figures 62 and 63.

It is also well known that a non-delayed talker sidetone has some influence on the subjective effects of talker echo. Figure 2 of the ITU-T Recommendation G.131 [14] for permissible talker echo, in principle the so-called 1 % curve, applies for "normal" values of STMR. If the STMR value is lower (<5 dB), the talker echo can be masked to a certain extent. Thus, an otherwise somewhat objectionable echo of moderate delay is subjectively "suppressed" by the strong sidetone (see subclause 9.2.5). On the other hand, if the STMR value is much higher (>20 dB or >25 dB), the talker echo becomes more noticeable. (For further discussions, see references [10], [41] and [43]. The mutual interaction effects between talker sidetone and talker echo can be estimated, at least approximately, by using the ETSI computation model described in subclause 9. Examples are shown in figures F.6 and F.7).

6.5 Echo and stability

6.5.1 General

Among others, impairments of speech quality may be caused by transmission delay and by echo effects, in conjunction with sources for coupling between the go and return path within a telephone connection. These effects are known as "talker echo" and "listener echo", depending, whether the talker is affected by his own reflected and delayed signal, or the listener observes additional (multiple) echoes of the direct signal.

Based on the same sources for coupling, singing or near singing distortion may arise within closed 4-wire loops, under specific conditions for the termination at the 2-wire ports (e.g. idle, short circuit). This attribute of closed 4-wire loops is called "stability".

The following paragraphs contain further information, definitions and acceptable limits and rules for echo and stability.

6.5.2 The principle of echo

6.5.2.1 Talker echo

The effect of echo is illustrated in figure 18. The transmitted speech signal (direct signal) of the talking subscriber is delayed along the different sections of the transmission path, coupled at the far end and received again with further delay, affecting the talker with an unwanted signal, comparable with an echo of his own voice. Since this type of echo in the given configuration is only observed by the talker, it is called "talker echo".

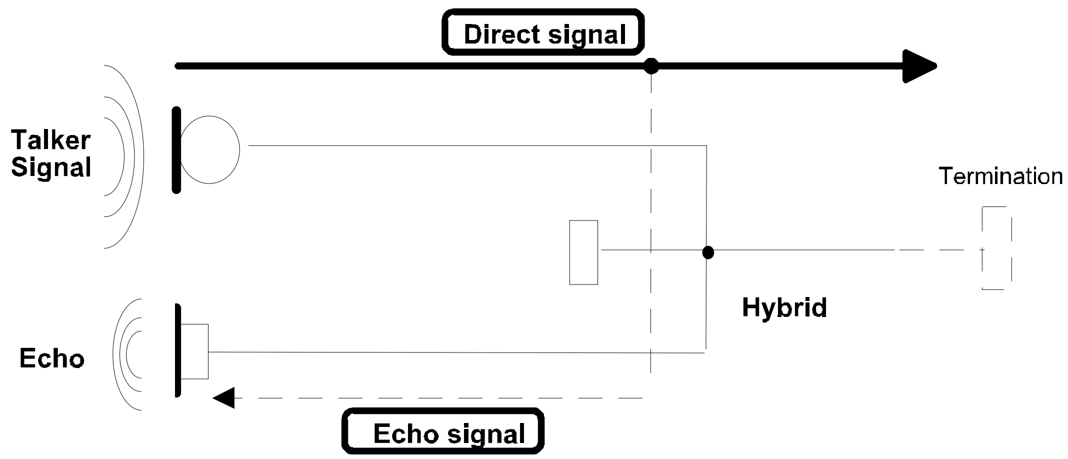


Figure 18: Effect of talker echo

6.5.2.2 Listener echo

In contrary to talker echo, if a closed 4-wire loop exists in a telephone connection as shown in figure 19, the talker's signal is coupled not only at the listener's end (hybrid B), but also again at the talker's end (hybrid A), therefore a double coupled signal will arrive additionally at the listener's end, some times later than the original signal. This delayed signal is called "listener echo". The main distinction from talker echo with only a single coupling is the double coupled signal, which may lead to an objectionable "hollowness" mainly in conjunction with low values for the transmission delay of only a few milliseconds.

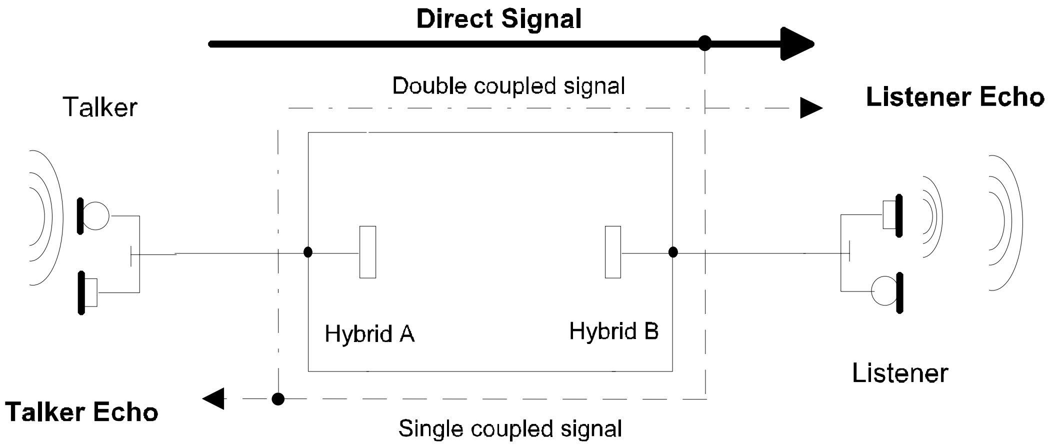


Figure 19: Effect of listener echo

6.5.2.3 Echo sources and echo loss

Sources for a coupling between go and return path causing echo effects in a speech conversation are mainly equipment, used for conversion between 4-wire and 2-wire, called hybrids. The amount of coupling of those hybrids depends mainly on the mismatch between the balance network Z_B and the terminating impedance Z_T at the 2-wire side (figure 20). The degree of mismatch is expressed as balance return loss a_{BRL} giving a direct impression about the magnitude of the coupled signal.

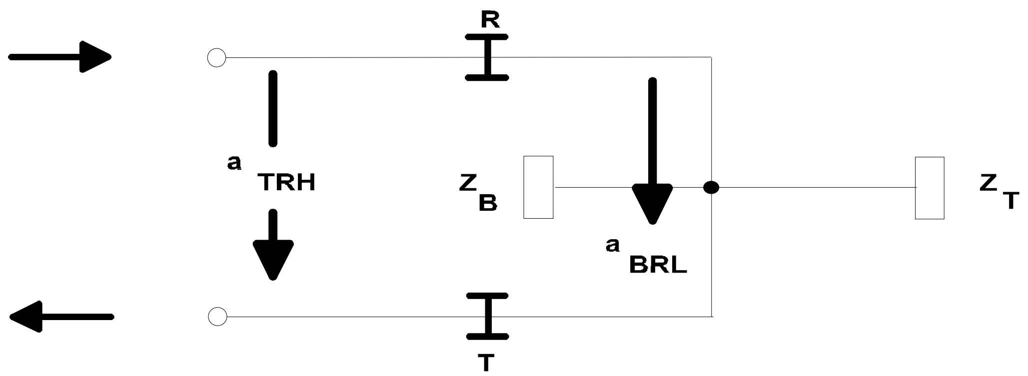


Figure 20: Balance return loss and transhybrid loss

Depending on the application within a network, hybrids also provide a specific loss in send direction (T-pad) and receive direction (R-pad) between the 2-wire and 4-wire ends, as shown in figure 20. The resulting loss between the 4-wire input and output, including the balance return loss and the loss of the R-and T-pads, is called transhybrid loss a_{TRH}.

Signal coupling may also arise at the interconnection of two 2-wire sections, such as switching equipment interfaces, transmission systems, cable sections and terminal equipment, if a mismatch between the two impedances exists. However in practice, network planning provides in most cases sufficient matching, therefore those coupling points are negligible for the effect of echo.

A further source for coupling exists within telephone sets, mainly in its acoustic path as shown in figure 21. Those sources are of increasing importance, since digital telephone sets will provide the only echo source in fully digital connections.

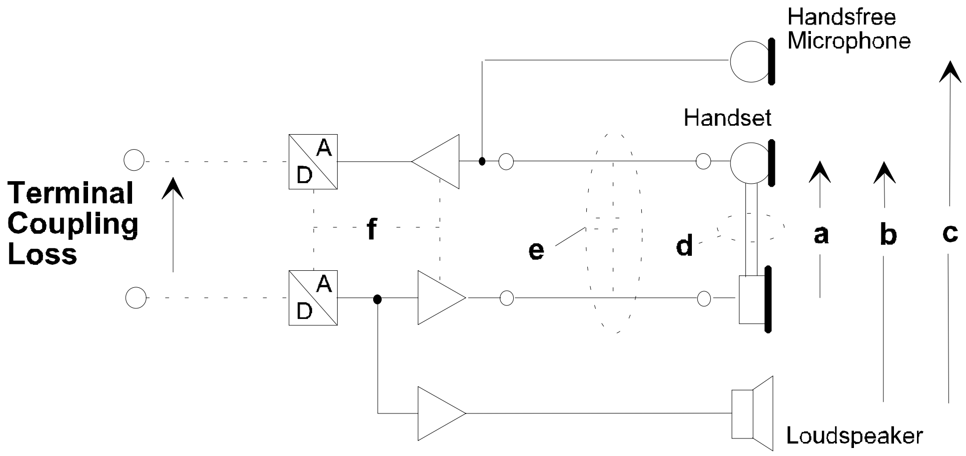


Figure 21: Coupling in a digital telephone set

In a digital telephone sets several coupling paths can be identified. Primarily the acoustic paths between receiver and transmitter capsule in the handset mode (path a in figure 21) and via loudspeaker and handsfree microphone in the handsfree mode (path c in figure 21) must be considered. In the loudspeaking mode both paths a and b (the path between loudspeaker and handset microphone) are active at the same time. Other paths such as mechanical coupling within the handset (path d in figure 21), capacitive coupling between the wires of the handset cord (path e in figure 21) and coupling via the power supply for codec and amplifiers (path f in figure 21) may have additional influence. All possible coupling paths are summarised and expressed in the term "terminal coupling loss" TCL, referred to the digital input and output of a digital telephone set.

In principle the same coupling paths with the exception of path f are given in an analogue telephone set with an analogue 2-wire interface. In those equipment, the different coupling paths will have a reaction to the input impedance as a source for return signals due to mismatch.

Coupling via hybrids or acoustic paths of telephone sets, are normally subject to an extensive shape in its frequency response. For the effect of echo, when considering the echo behaviour of a hybrid, the transhybrid loss is weighted with a specific weighting scale over the frequency range 300 Hz to 3 400 Hz. This weighted transhybrid loss is then called echo loss. For digital telephone sets presently the same weighting is used and expressed as weighted terminal coupling loss TCLw.

According to ITU-T Recommendation G.122 [25] section 4.2 echo loss EL and TCLw are derived from the integral of the power transfer characteristic A(f) weighted by a negative slope of 3 dB/octave from 300 Hz to 3 400 Hz as follows:

$$\mathrm{EL} = 3{,}85 - 10.\mathrm{l\,g}\left\{ \int_{300}^{3\,400} . \frac{\mathrm{A}(f)}{f} df \right\} \quad (\mathrm{dB}) \qquad (6.5.1)$$

where

$$\mathrm{A}(f) = 10^{-Lab(f)/10} \qquad (6.5.2)$$

with L_{ab} as the loss of the echo path at frequency *f*. If the results are available in graphical form or as tabulated data, the echo loss may also be calculated using the trapezoidal rule. More information are given in Annex B to ITU-T Recommendation G.122 [25], section 4.

NOTE: In USA a weighting algorithm according to an ANSI-Standard is used. The use of a weighting according to Annex B of ITU-T Recommendation G.122 [25], section 4 for TCLw with its acoustical coupling needs further study (see also Annex M, clause M.2).

The magnitude of the talker echo is characterized by the talker echo loudness rating, TELR.

TELR = EL + SLR + RLR

Here SLR and RLR are the send and receive loudness ratings of the talker's telephone set, referred to that 4-wire interface where the echo loss EL applies.

6.5.2.4 Multiple echoes

Figures 18 and 19 demonstrate the echo effect using equipment with a 4-wire to 2-wire conversion (hybrids), as they are mainly part of analogue or mixed analogue/digital transmission paths in a connection. Those mixed connections may in many cases consist of more than one 4-wire loop. Figure 22 shows the configuration for talker echo for a connection via two 4-wire loops. In this case 3 different talker echo paths exist, including the acoustic path at the listeners end. Those configurations result in a summation of several multiple echoes at the talkers end and some more echo paths for the listener echo. In practice, usually each of the different echo paths will be investigated separately and the most critical one, with respect to its influence to the MOS-rating, will be the basis for planning. This can be done using the computation model for transmission quality (see clause 9), computing the MOS-rating for each loop, using the transmission delay and TELR of the considered echo path, assuming no coupling in the remaining loops.

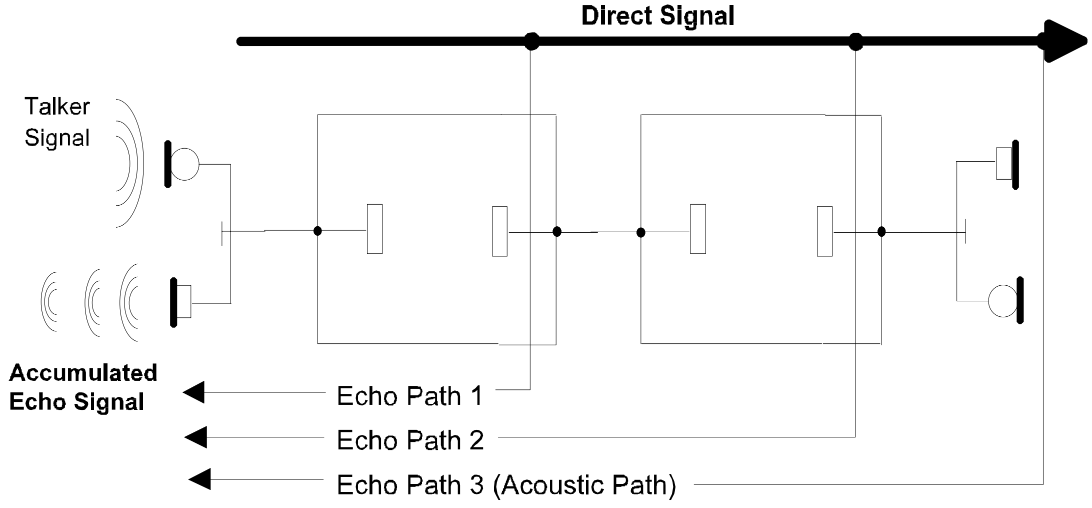


Figure 22: Multiple talker echo

6.5.3 Stability

Beside the effect of talker echo and listener echo caused by coupling between go and return path, two hybrids forming a closed 4-wire loop may also, under specific conditions, result in an oscillation within the loop, called "singing". The 4-wire loop in an mixed analogue and digital international or national connection, provides gains "s" and losses "a" including the balance return losses "a_{BRL}" at the two terminating hybrids as shown in figure 23. The point where singing starts is reached, when the sum of all losses and gains is equal or less than 0 dB.

The attribute of a 4-wire loop with respect to a possible singing is called its stability. The sum of all losses and gains within the loop, responsible for the stability is called the open loop loss OLL. If this OLL is close to the singing point of 0 dB, a further impairment of the speech quality may arise due to the "near singing distortion". Therefore a stability margin is required to avoid those effects.

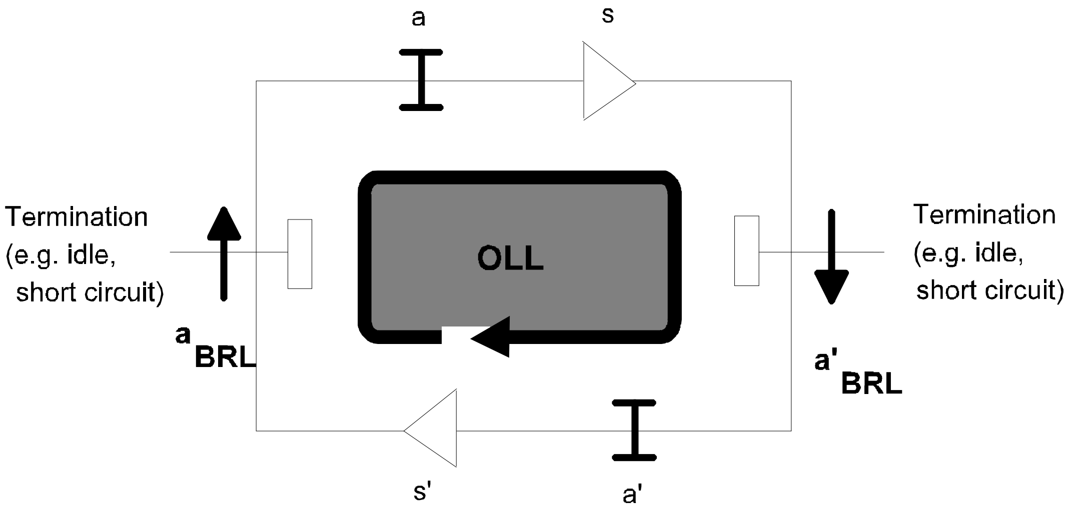


Figure 23: Stability of a 4-wire loop

In contrary to talker echo and listener echo, where only frequencies in the range from 300 Hz to 3 400 Hz in conjunction with an echo weighting are considered, a wider range of 0 Hz to 4 000 Hz must be taken into account for singing, since oscillation will start at every single frequency for which the OLL has its minimum.

A sufficient stability margin must be provided in any case, to avoid overloading of systems and crosstalk into other channels of FDM and cable sections.

In configurations, where the OLL for any of the individual single 4-wire loops is near to the stability limit, the OLL for each compound 4-wire loop must be higher than for a single loop to avoid problems of interaction of the single loops.

6.5.4 Definitions for echo and stability

In conjunction with echo and stability the following terms and definitions are used.

Echo generally is defined as an unwanted signal, delayed to such a degree that it is perceived distinctly from the wanted signal (direct transmitted signal). Echo is subdivided into "talker echo", where the coupling occurs near the listeners end, affecting the talker and the "listener echo", defined as an echo, produced by double coupled signals and disturbing the listener.

The echo path is understood as the transmission path, via the send and receive direction of the national and international chain and the terminating point where the coupling occurs.

Mainly for the consideration of echo effects the "echo path delay" is important. It expresses the transmission delay along the echo path in milliseconds. Transmission delay is not only caused by the propagation delay along transmission media, but also due to delay mainly in digital systems and digital signal processing.

More commonly in use is the term "mean one-way transmission time", which is defined as half the sum of the transmission time in both transmission directions.

When considering signal couplings, the magnitude of these coupling is generally depending on the mismatch between two impedances Z1 and Z2 and expressed as "return loss" L_R. The definition for return loss is given by the equation

$$L_R = 20\lg\left|\frac{Z1+Z2}{Z1-Z2}\right| \text{ dB} \tag{6.5.3}$$

When hybrids as a source for signal coupling are considered, the relevant impedances are the balance impedance ZB, the input impedance Z0 of the hybrid at the 2-wire side and the impedance Z2 terminating the 2-wire side of the hybrid. This mismatch is characterized by the term "balance return loss" L_{BR} and given by the formula

$$L_{BR} = 20\lg\left|\frac{Z0+ZB}{2Z0}\,x\,\frac{Z2+Z0}{Z2-ZB}\right| \text{ dB} \tag{6.5.4}$$

In many cases, the input nominal impedance Z0 is equal to the nominal terminating impedance Z2, and the formula for L_{BR} is reduced to

$$L_{BR} = 20\lg\left|\frac{Z2+ZB}{Z2-ZB}\right| \text{ dB} \tag{6.5.5}$$

With respect to echo, the balance return loss is averaged with 1/f power weighting over the telephone band and called "echo balance return loss". It forms the "echo loss" of the echo path together with the sum of the transmission losses in both transmission directions.

The loss of a path comprising one or more 4-wire circuits and a 4-wire/2-wire terminating set, is also called "semi-loop loss". It may either be considered for a given piece of equipment, or for the national system of an international connection. In the latter case the echo loss is identical with the weighted a-t-b loss as shown in figure 24. In this semi loop the 4-wire input and output are referred to the international connecting point ICP.

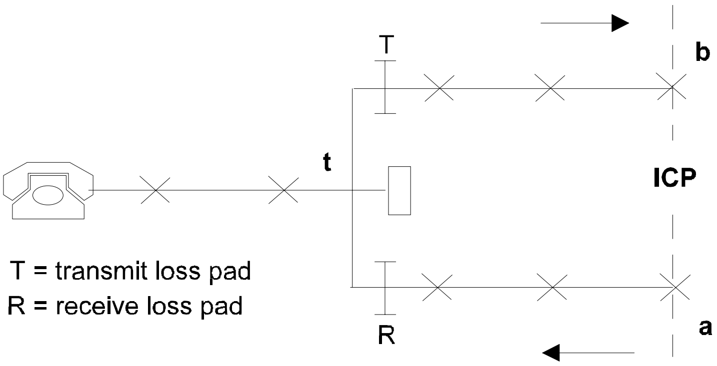


Figure 24: Loss of the path a-t-b of a national system

The echo path may also be terminated by a 4-wire (digital) telephone set. With respect to echo, those terminals are characterized by the parameter "weighted terminal coupling loss" TCLw, including acoustical coupling at the users interface, electrical coupling due to crosstalk within the handset cord or electrical circuit and mechanical coupling through the mechanical parts of the terminal.

Beside the mean one-way transmission time the second important factor for an objectionable echo is the "talker echo loudness rating" TELR. It is defined as the sum of the sending loudness rating (SLR), the receiving loudness rating (RLR) of the talkers telephone set and the echo loss of the echo path. For the consideration of international connections it is usually defined as the SLR and RLR of the talker's national system, twice the loss of the international chain and the echo loss (semi-loop a-t-b) of the listener's national system.

For listener echo similar terms are used. The "listener echo loss" LE is defined as the degree of attenuation of the double coupled signal L_2, with respect to the wanted signal L_1.

$$LE = L_2 - L_1$$

For practical purposes the LE is equal to the "open loop loss" OLL, the sum of all losses and gains including the echo balance return losses within the 4-wire loop. The LE characterizes also the degree of disturbance by hollowness, an effect subjectively perceived as a "hollow sound" i.e. as if the talker would speak into a some hollow vessel.

The level difference between the speaker's direct voice and the listener echo is the same as a weighted mean value of OLL over the speech band. This mean value is also to be interpreted as a special loudness rating and is termed "Weighted Echo Path Loss" (WEPL).

According to ITU-T Recommendation G.126 [116], the expression for WEPL is

$$WEPL = -20.\lg\left[\frac{1}{3\,200}\int_{200}^{3\,400} 10^{-0{,}05.OLL(f)}\,df\right] \tag{6.5.6}$$

Note that this flat weighting is done on a linear frequency scale and is based on voltage addition. (However, in practice almost the same result will often be obtained with the weighting done on a logarithmic frequency scale. Also, with sufficiently good accuracy, WEPL can be calculated quite simply as the sum of the weigted means of the two semi-loop losses and with the weighting based on a power basis, as is done for the talker echo).

Just as for talker echo, the subjective annoyance of the listener echo increases with longer delays. (An estimation can be made by using the ETSI computation model, see clause 9). However, if the talker echo is under control, the listener echo usually does not cause problems.

The "stability loss" can be defined using the same configuration as for echo loss (semi-loop loss) or terminal coupling loss. The stability loss is the lowest loss (without weighting) within the considered frequency band and configuration.

6.5.5 Methods for the control of echo and stability

The following subclauses describe some methods, applicable to keep the margin for stability and echo within the limits given in subclause 6.5.6. For the stability of the 4-wire chain of a connection and for the control of echo in connections with low values of transmission delay, mainly the loss contribution of the connection is used to provide the necessary performance. For connections with higher values of transmission delay, specific equipment, called "Echo Control Devices (ECD)" such as echo suppressors or echo cancellers are used.

6.5.5.1 Loss contribution

As specified in subclause 6.5.6 a possible way to control stability and - for low values of transmission time - echo effects, is the provision of loss for the considered path. Since hybrids as the main means of coupling in analogue or mixed analogue/digital connections are responsible for both stability and echo performance, the balance return loss and the resulting transhybrid loss are of major importance. Additional losses in the 4-wire chain may support the loss contribution, however, in case of analogue transmission systems like FDM also loss deviations from its nominal value must be considered.

The provision of loss is limited by the Overall Loudness Rating (OLR) of a connection. Therefore loss contribution, mainly in an analogue or mixed analogue/digital network must be in any case a compromise between the requirements for OLR and echo or stability. In most of all cases the long distance section in national and international connections consists of 4-wire chains with terminating 4-wire/2-wire conversions in the lower network hierarchy. These 4-wire loops are usually adjusted to a nominal loss of about 7 dB between the 2-wire ends. The resulting minimum OLL of 14 dB, is only reduced by loss deviations contributed from analogue FDM systems. However, the introduction of digital switching systems in local exchanges or PBXs with 2-wire interfaces is resulting in additional 4-wire loops for connections in mixed analogue/digital networks. Due to national transmission planning the insertion loss of those exchanges should normally be low, in the range of 1 - 2 dB only. Therefore care must be taken for the design of the hybrids and the choice of balance networks matching to the impedance of the connected 2-wire sections.

In connections with fully digital routing and switching, terminated only in a hybrid at one or both ends, the balance return loss in conjunction with R- and T-pads is usually the remaining source to provide the necessary loss for the control of echo and stability. Additional inserted digital pads may increase the loss, but contribute with quantization distortion and a reduction of the dynamic range.

For a closed 4-wire loop, the OLL provides the necessary singing margin. The most critical point for singing is during call set-up and release of a connection. In these phases, idle or short circuit termination at the 2-wire side of the hybrids will exist, so the balance return loss will decrease to values, such that a sufficient singing margin can no longer be provided. In those cases modern switching equipment do not establish the call path until the answer signal is received and a correct termination at the 2-wire side is provided.

The detection of echo signals and the subjective judgement is mainly based on the volume of the received echo signal and the delay. In practice, the volume of the echo signal is expressed by the "talker echo loudness rating" TELR, which is equal the sum of the echo loss along the echo path and the SLR and RLR of the used telephone set at the talker's end. Also in this case the requirement for a high echo path loss is limited by the OLR.

According to subjective tests, the TELR should increase with the mean one-way transmission time. As a recognized rule (see also subclause 6.5.6) values up to 25 ms can be controlled by loss contribution. For higher values echo control devices must be used.

6.5.5.2 Echo control devices

Subclause 7.1 describes more in detail the principle of Echo Suppressors (ES)s and Echo Cancellers (EC)s presently in use. Further information is also given in annex D, and in CCITT Recommendations G.164 [18] for ESs and G.165 [19] for ECs.

Presently, echo control devices are used for suppressing echo in connections with a mean one-way transmission time of more than 25 ms, i.e. mainly in international connections. The basic application of ECD is shown in figure 25. At both ends of an established connection an ECD is inserted, usually in International Switching Centers (ISCs). The echo path is in these applications the remaining national part of the call path.

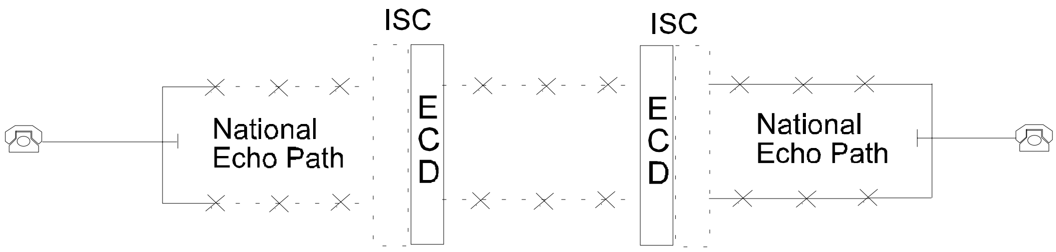


Figure 25: Application of echo control devices

The main task of an ECD is to suppress or compensate signals arising from a coupling between the go and return path (echo path) without inserting other remarkable impairments to the speech quality. Two main basic principles are known and in use for several years, the echo suppressor and the more modern one, the echo canceller.

The national echo path, mainly the echo path delay, depends on the geographical size, location of ISCs and routing in a national public network. For average sized countries, maximum mean one-way transmission times in the range of 15 ms can be expected, leading to echo path delays in the range of 30 ms to 50 ms. Echo suppressors are only applicable for echo path delays up to 25 ms. On the contrary, echo cancellers are typically designed to compensate echo path delays in the range from 45 ms to 60 ms.

Common to all ECDs is a minimum necessary echo path loss of 6 dB for proper operation. As far as ESs and ECs are designed according to CCITT Recommendations G.164 [18] and G.165 [19], compatibility is guaranteed also between an ES inserted at the one end and an EC at the other end.

6.5.6 Acceptable limits and planning rules

6.5.6.1 Limits for stability

For the commitment of values for stability loss, a guidance is available in ITU-T Recommendation G.122 [25]. However these recommended values are referred to the path a-t-b (see figure 24) as involved in international connections. It is recommended that the sum of the nominal losses a-t and t-b is equal or greater than 4 dB + n, where n is the number of 4-wire analogue or mixed analogue/digital circuits in the national section. Assuming furthermore a minimum balance return loss at the terminating hybrids during normal operating conditions of 2 dB, the recommendation results in a stability loss of

Mean value = $6\mathrm{dB} + \mathrm{n}$

Standard Deviation = $\sqrt{4\mathrm{n}}$ dB

With the increasing use of fully digital circuits, the required stability loss is 6 dB. This value is applicable not only to 4-wire/2-wire conversions located within public network, e.g. in digital local exchanges, but also to hybrids in digital PBXs, digitally connected to the public network and terminating an international connection. With respect to the ICP those semi-loop losses contain additionally R- and T-pads with a sum in the range of 5 dB to 7 dB, so the required stability loss of 6 dB is guaranteed.

A sufficient stability loss must not only be provided for the national or international 4-wire chain, but also for every 4-wire loop as they are formed by digital local exchanges or digital PBXs in a 2-wire environment. The issue of a minimum stability loss for each of the two semi-loops referred to a virtual reference point in the center of the exchange (switching matrix) is no longer applicable.

The loss allocation in those exchanges is usually, due to other non-symmetrical reasons, i.e. one of the two semi loops may provide a sufficiently high loss, while the second one contributes with a gain to meet the requirements for a low value of insertion loss. The provision of a defined stability loss can only be described in a minimum OLL for the considered configuration. A suitable nominal value could be

$$\text{OLL} \geq 4 \text{ dB}$$

assuming normal operating conditions, negative deviations due to frequency response in the range from 300 Hz to 3 400 Hz are limited to about 0,5 dB and loss variations with time are negligible. The loss below 300 Hz and above 3 400 Hz will be in any case higher than within the speech band, due to the coding/decoding process and sufficient filtering.

The value of OLL 4 dB is a sufficient value also with respect to possible near singing distortions. However, different facts must be taken into account. With respect to capacitive complex balance networks and differences in the values of balance impedance and input impedance at the 2-wire side of the hybrid, the balance return loss may move towards 0 dB and in critical configurations also to small negative values. To meet in any case the requirement of 4 dB, the nominal insertion loss must be adjusted in the range of 2 dB to 2,5 dB, which may exceed the limits given at present by the national loss planning mainly for PBXs.

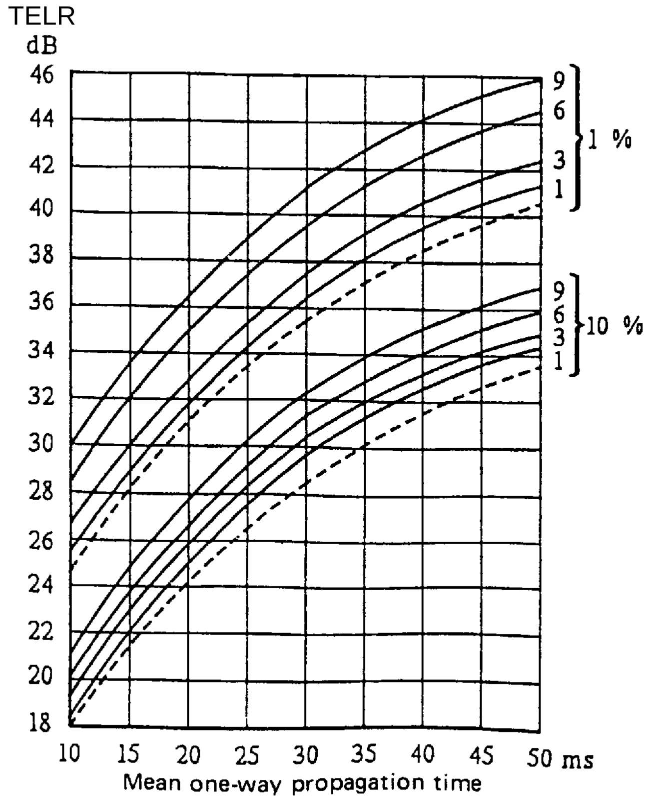
NOTE: In some switching exchanges during call set-up and release phases, the switching matrix is not through connected until a correct answer signal is received and a sufficient termination at the 2-wire ends is provided.

6.5.6.2 Limits for echo

Basically, the impact on speech quality by echo, mainly the talker echo, increases with the volume of the received echo and the echo path delay. The degree of impairment must be established by subjective tests, varying the mean one-way transmission time and the echo path loss. As a simple rule, degradation of speech quality by echo effects can be avoided, if the necessary echo path loss is provided for every value of transmission time.

Several subjective tests have been carried out, to determine the echo loss / transmission time ratio based on a MOS rating about the speech performance, or for the objection of echo. Based on a series of studies by AT&T completed in 1971, echo tolerance curves has been derived for the overall loudness rating via the mean one-way transmission time. These curves, published in figure 2 of ITU-T Recommendation [14], are in common use for network planning by almost all administrations and network operators since several years. They must be interpreted in that way, to obtain information whether, if in a given configuration of OLR and transmission time, additional ECDs are required.

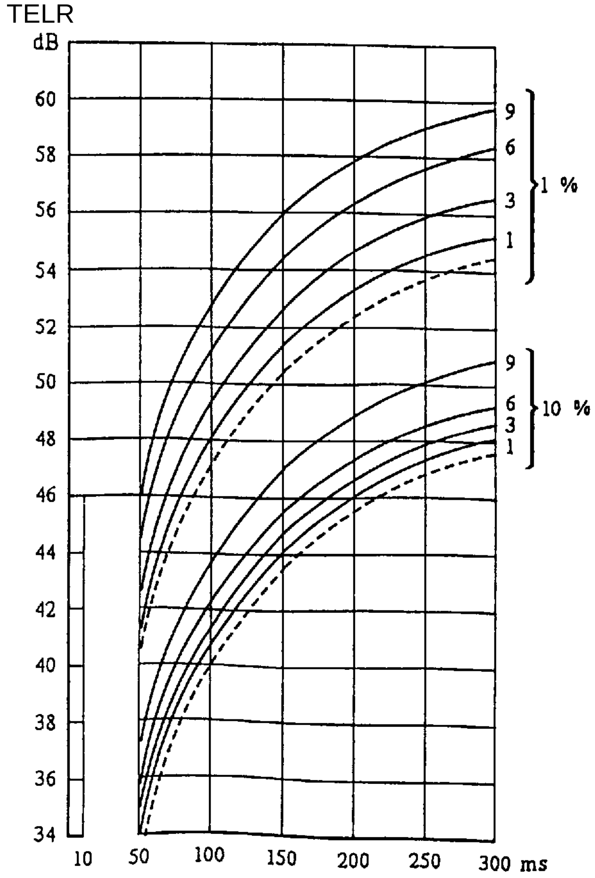
The curves, reproduced here as figures 26 and 27, apply for fully digital connections and for different numbers of analogue 4-wire circuits, for mean one-way transmission time in the range from 10 ms to 300 ms and for 1 % or 10 % probability of encountering objectionable echo. The values for 10 % probability should only be used in exceptional cases. (Note that the term OLR in figure 2 of ITU-T Recommendation G.131 [14] is identical with the talker echo loudness rating TELR as used in this ETR and defined in subclause 6.5.4).



——— Analogue circuits, with indication of number of 4-wire circuits

- - - - - Fully digital connections

Figure 26: Echo tolerance curves, range 10 - 50 ms. The percentages refer to the probability of encountering objectionable echo



——— Analogue circuits, with indication of number of 4-wire circuits

- - - - - Fully digital circuits

Figure 27: Echo tolerance curves, range 50 - 300 ms. The percentages refer to the probability of encountering objectionable echo

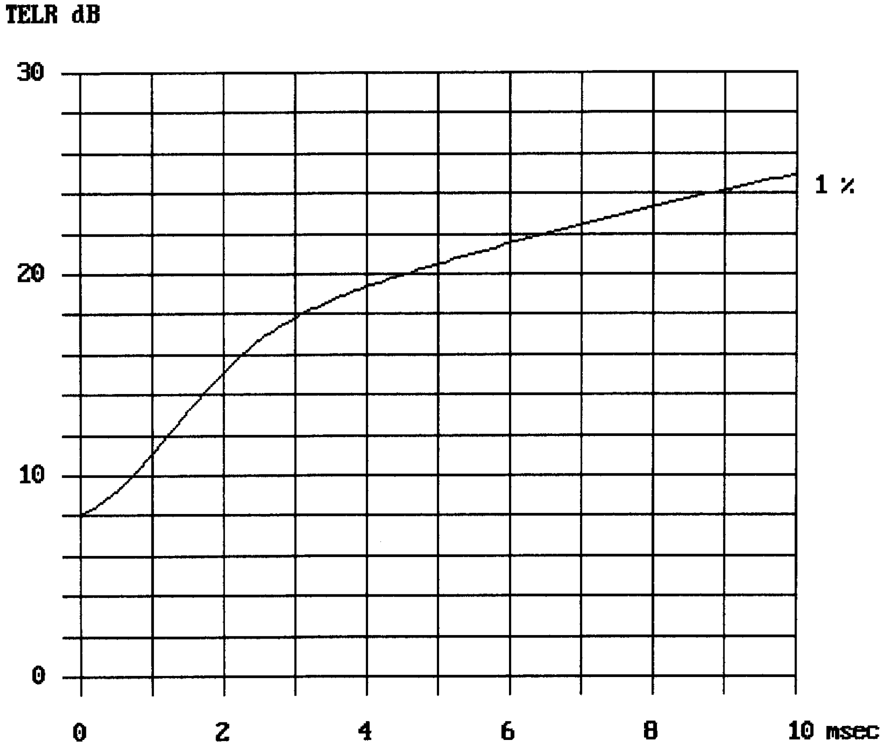
With a good approximation these curves can also be expressed by the following equation:

$$TELR = A0 - 6e^{-0,3T^2} + 40\lg\frac{1+T/10}{1+T/150} \qquad (6.5.7)$$

where:

T = Mean one-way transmission time in ms;
A0 = 14 dB for 1 % objection, 7 dB for 10 % objection.

Note that equ.(6.5.7) also holds for values of *T* lower than 10 ms, as has been shown by comparisons with subjective tests results from France Telecom and Telia Research. Figure 28 shows the computed "1 % curve" for TELR as function of T in the range 0 ms to 10 ms. (Note, however, that for T < 1 ms, TELR is to be interpreted as STMR for which the permissible range is much wider, see subclause 6.4).



NOTE: The percentage (1 %) refers to the probability of encountering objectionable echo, mainly in the range 1 - 10 ms. For delays in the order of 1 ms and lower, the talker echo is perceived more as a talker sidetone.

Figure 28: Echo tolerance curve, range 0 - 10 ms, as computed by equ.(6.5.7)

As an example, if a fully digital connection is assumed for a national call with 25 ms one-way transmission time, for 1 % probability a TELR of 33,1 dB is necessary to avoid the use of ECDs. This formula may also be used for values of transmission times below 10 ms, a range which is not covered by the curves in ITU-T Recommendation G.131 [14]. For very low values of transmission time the resulting TELR will be interpreted as STMR. For T = 0 ms the required STMR will be 8 dB for 1 % probability. It should be noted however, that echo loss and STMR use different weighting algorithms.

There is less information available up to now about degradation of speech quality which can be considered as a "delayed sidetone" in this range below 10 ms and a possible masking of echo effects by the STMR of the telephone system. Mainly in private networks with internal digital routing but analogue 2-wire connections to the public network, such delayed sidetone may cause an impairment of transmission performance.

EXAMPLE: Assuming a digital telephone set in a private network, connected digitally to the 2-wire analogue trunk-interface with the public network as shown in figure 29, with 5 ms one-way transmission time and a SLR of +3 dB and a RLR of -2 dB (values to meet national loss planning requirements for analogue access), the necessary echo path loss can be calculated as follows:

From the equation above, with 1 % probability the required TELR is calculated with 20,5 dB. Since the sum of SLR and RLR is +1 dB, the required echo path loss is 19,5 dB. Taken into account that the only loss within the considered echo path is provided by the transhybrid loss of the 4-wire to 2-wire conversion of the trunk interface, which has a loss adjustment resulting in a gain of 3 dB, the required balance return loss is 22,5 dB. In practice, those high values cannot be guaranteed, mainly with a fixed balance impedance but varying input impedance to the public network, depending on the type of connection and routing.

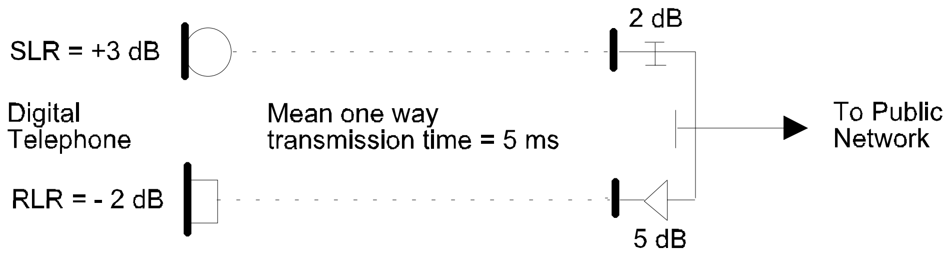


Figure 29: Example for TELR in a private network

If the delay is reduced to a more practical value of only 1 ms, as in a single stand alone digital PBX, the required TELR is reduced to 11,1 dB and the balance return loss in the given example to 14,1 dB, a value which may not be met in every practical configuration, but there are not reported troubles from customers using several thousands of comparable PBX configurations.

Using the equation above for the maximum permitted transmission time of T = 400 ms, the resulting TELR is 56 dB. For digital telephones, designed to have a nominal SLR = +7 dB and RLR = +3 dB, the echo path loss for 1 % echo objection must be 46 dB. Assuming fully digital connections with incorporated digital telephone sets on both sides and a handset which provides a TCLw of more than 46 dB, no echo control devices need to be inserted.

As far as analogue telephone sets terminate such a call, the same values for transhybrid losses are impossible to realize, i.e. ECDs must be used. The same rule is valid for calls via analogue or mixed digital/analogue circuits, since those sections contributing with the main part of transmission time are usually national or international 4-wire circuits with nominal losses of 0 dB or 0,5 dB only.

Planning rules for the control of echo should therefore be based on the value of 25 ms for the mean one-way transmission time of a connection. The required TELR for 1 % probability is in this case 33 dB, which can be established also in fully digital connections without any other impairment to transmission performance. This enables most of the European network operators to avoid the use of ECDs in national calls.

Planning rules must also include private networks and the allocation of transmission time. If established values for the delay are exceeded by those networks or by mobile/cordless terminal equipment, then those terminations must incorporate echo cancellers. In those cases two pairs of ECs may be inserted in an international connection. Practical experiences with respect to connections including ECs in tandem configuration, performed by AT&T, indicate that there are no significant differences in the MOS rating between single or double pair of echo cancellers. However this item is under further study. For further information see also subclause 7.1 of this ETR.

6.6 Transmission time

The task to derive acceptable limits for transmission time and echo in an established connection must consider two different influences to the overall quality of speech transmission. Both types of impact on speech quality are caused by the physical effect of propagation time via the different media used in a telephone connection. The increasing implementation of digital technology is resulting in additional delay due to digital signal processing, use of low bit-rate coding, and processing time in digital switching centers.

In conjunction with signal coupling, transmission time will cause echo (see subclause 6.5), but also in echo free connections when echo control devices are used or the echo loss is sufficiently high, the total transmission time must be limited.

6.6.1 Limits for transmission time

If the mean one-way transmission time increases to about some hundreds of milliseconds, problems may arise to the talkers to follow an interactive normal flow of conversation and to interrupt. Recent investigations regarding long pure delay, published in Annex B of the revised ITU-T Recommendation G.114 [36], provide some information about the result of subjective tests. As far as highly interactive talks are considered, combined with specific measures to difficulties such as ability to interrupt, the effect can be detected well below a value of 400 ms.

The improvement of echo control devices, mainly echo cancellers, in the past years however has extended the pure delay values with a poor or worse rating compared with earlier investigations. Assuming the increasing use of high performance echo cancellers and a careful planning with respect to equipment causing additional delay, the following guidance about the precautions and limits for the transmission time can be given, derived mainly from the well experienced issues in ITU-T Recommendation G.114 [36]:

a) range 0 - 25 ms

This range of transmission time can be expected for national calls within average sized countries. There are no difficulties with respect to a normal flow of conversation. Usually echo is controlled by providing a sufficiently high echo loss instead of echo control devices;

b) range 25 - 150 ms

This range is acceptable for most user applications assuming the use of echo control devices;

c) range 150 - 400 ms

Within this range - in most cases including a satellite link - difficulties may arise for interruptability and normal flow of conversations, mainly in high interactive talks. High performance echo cancellers according to ITU-T Recommendation G.165 [19] must be used and careful network planning is necessary;

d) range above 400 ms

Values of transmission time above 400 ms must be avoided for general network planning.

NOTE: The value of 400 ms should be exceeded only in exceptional cases. Those exceptions are for example unavoidable double satellite hops, satellites used to restore terrestrial routes, one or more digital cellular interconnections and combinations of mobile systems and satellite links.

For planning purposes there is no guidance up to now about the allocation of total transmission time to the national and international sections of a connection. For international calls but also for some national calls an increasing use of satellite links can be expected. Assuming a satellite link as the only international part of a connection, with 260 ms between the two earth stations, the remaining maximum transmission time for each national section is 70 ms.

In most European countries - according to the geographical size - the maximum transmission time used by the public networks between the network connection points (NCP), will be in the range of 15 ms to 25 ms. This value can normally also be assumed for the section NCP to international switching center (ICP). Therefore, for terminal equipment or private networks connected to the public networks, a maximum value for the transmission time of 45 ms to 55 ms is available. This may allow the use of mobile or cordless systems for terminal equipment, or a limited insertion of multiplexing equipment within private networks.

However it should be noted, that in those cases the use of echo control devices is mandatory also for normal national calls. Furthermore those terminal equipment or private networks involved in an international connection may be subject to difficulties in conversation as described above.

6.7 Noise and quantizing distortion

6.7.1 General

Noise used to be a very important limiting factor for the speech quality of long-distance connections in the days of FDM (carrier) transmission systems. For digital systems noise is much less of a problem. However, noise in various forms still influences the customers' perception of the connection quality. Here a short survey will be given of how to consider the effects of noise. The following points will be discussed briefly:

- circuit noise;
- quantizing noise;

- addition of noise sources in a connection; equivalent noise in a 0 dBr point;
- ambient noise;
- induced noise on subscriber lines;
- the subjective effects of non-stationary noise.

A more comprehensive list of references for the subject can be found in annex B: "Noise aspects in understanding overall network performance".

6.7.2 Circuit noise, quantizing noise and distortion

In ITU-T Recommendation G.222 [20] a noise objective of 10 000 pW0p or -50 dBm0p is recommended for the design of carrier systems of 2 500 km.

CCITT Recommendation G.143 [117] states that the total noise generated by a chain of six international circuits should not exceed -43 dBm0p when referred to the first circuit in the chain. The limit for a single tone noise should be 10 dB lower than the psophometric noise power in the circuit and, to avoid audibility, an additional margin of 5 dB is recommended. The limits for narrow bands of noise should be more stringent than those for wideband noise.

According to ITU-T Recommendation G.712 [24], for a 64 kbit/s PCM codec pair, the idle channel noise should not exceed -65 dBm0p. The level of any single frequency (in particular the sampling frequency and its multiples), measured selectively, should not exceed -50 dBm0p. Noise contributed by the receiving equipment alone should be less than -75 dBm0p when its input is driven by a PCM signal corresponding to the decoder output value number 0 for the µ-law or decoder output value number 1 for the A-law.

Quantizing distortion appears as a signal-correlated noise. For PCM systems it is quantified by the "quantizing distortion unit" (qdu). One qdu is defined as the degradation given by an average 8-bit codec pair which complies with CCITT Recommendation G.711 [62] (see ITU-T Recommendation G.113, clause 3.1 [35]). (Note that an average codec pair produces about 2 dB less quantizing distortion than the limits indicated in ITU-T Recommendation G.712 [24]). Table 3 gives some planning values for qdus; source ITU-T Recommendation G.113 [35].

For telephone connections, which incorporate unintegrated digital processes, it is permissible to add the qdus that have been assigned to the individual digital processes.

Table 3: Planning values for quantizing distortion; qdu

Digital process	qdu
8-bit PCM codec pair (ITU-T Recommendation G.711 [62]), between analogue interfaces	1
32 kbit/s ADPCM (ITU-T Recommendations G.726 [121] and G.727 [66]) and 16 kbit/s LD-CELP (ITU-T Recommendation G.728 [81]), between analogue interfaces	3,5
Digital loss pad, additional qdu	0,7
PCM/ADPCM/PCM conversion (ITU-T Recommendations G.726 [121] and G.727 [66])	2,5

For low bit-rate codecs, the qdu concept is not quite suitable. For those, the use of "equipment impairment factors" as described in subclause 9.1.3.5 and annex G appears to be more suitable.

6.7.3 Addition of noise sources in a connection; equivalent noise in a 0 dBr point

According to ITU-T definitions, a "connection" is constituted by several "circuits" in tandem. A circuit usually consists of the permanently interconnected transmission links between two exchanges.

As the voice signal passes through a circuit, noise is added by various links in the circuit chain. The noise levels are conveniently expressed in dBm0p. At the end (output) of the circuit the total noise is:

$$Nc = 10 \cdot \lg\left\{\sum 10^{Ni/10}\right\} \quad \text{dBm0p} \qquad (6.7.1)$$

where *Ni* is the noise (in dBm0p) introduced by link number *i*, and $i = 0$ corresponds to the input noise fed to the circuit under consideration by the preceding circuit.

It is sometimes convenient to think of the total circuit noise of a circuit as being created by a single fictitious noise source of *Nc* dBm0p introduced at a 0 dBr point, which can be termed "the equivalent noise in a 0 dBr point".

If the relative level at the output of a circuit is *Lr* dBr, the absolute noise level at the output is

$$Na = Nc + Lr \qquad \text{dBmp} \qquad (6.7.2)$$

At the interconnection point of two circuits, sometimes the output relative level of the sending circuit is lower by *S* dB than the input relative level of the receiving circuit, a "level jump". The input noise (to the receiving circuit) in dBm0p is obtained by subtracting *S* dB from the output noise (of the sending circuit) in dBm0p. (Level jumps are introduced for stability reasons in analogue and mixed digital/analogue connections. Usually $S = 0{,}5$ dB).

6.7.4 Ambient (room) noise

The disturbing effect of ambient noise in a location is usually characterised by its value in dB(A) which is a frequency-weighted power average that can be measured by special sound level meters. Table 4 lists examples of typical dB(A) - values for some ambient sound conditions.

Ambient (room) noise can be picked up by the telephone set microphones both at the send and receive side of a connection and transformed into equivalent circuit noise. The governing parameters are:

- room noise at the send side: *Pos* dB(A);
- room noise at the receive side: *Por* dB(A);
- send Loudness Rating, referred to the 0 dBr point: SLR;
- receive Loudness Rating, referred to the 0 dBr point: RLR;
- the handset direct-diffuse sensitivity factor, send side: Ds;
- the Listener Sidetone Rating LSTR, receive side: LSTR.

NOTE: The factor D is the difference in (weighted) average handset sensitivity in dB between direct (voice) sounds and diffuse (room) sounds. D is in the order of 3 dB for linear microphones in conventionally shaped handsets. For unsuitably shaped, short handsets D = 0 or even less. With carbon microphones, the non-linearity makes D to be about 6 dB.

The equivalent circuit noise *Nos,* referred to a 0 dBr point, caused by room noise *Pos* dB(A) at the send side, is:

$$Nos = Pos - SLR - Ds - 100 + 0{,}008\left(Pos - OLR - Ds - 14\right)^2 \quad \text{dBm0p} \qquad (6.7.3)$$

where

$$OLR = SLR + RLR \qquad (6.7.4)$$

The equivalent circuit noise *Nor,* referred to a 0 dBr point, caused by room noise *Por* dB(A) at the receive side, is:

$$Nor = RLR - 121 + Pore + 0{,}008\left(Pore - 35\right)^2 \quad \text{dBm0p} \qquad (6.7.5)$$

where *Pore* is the *effective* room noise caused by enhancement of *Por* by the listener's sidetone path

$$Pore = Por + 10\lg\left[1 + 10^{(10 - LSTR)/10}\right] \qquad \text{dBm0p} \qquad (6.7.6)$$

These noise sources are added to the other circuit noises as described by equ.(6.7.1).

Table 4: Examples of sound (noise) conditions

Examples of sound (noise) conditions	dB(A)	dBPa(A)
Hearing threshold	0	-94
Lower limit for speech to be intelligible	15	-79
Very quiet office	20	-74
Office with forced ventilation	40	-54
Noise level in a bedroom above which sleeping becomes difficult	40	-54
Highest in-door noise level that does not cause discomfort	45	-49
Highest out-door noise level that does not cause discomfort	55	-39
Normal conversation	55	-39
Heavy traffic	70 ... 80	-24 ... -14
Ambient noise limit above which a telephone conversation is difficult	70	-24
Pneumatic air drill	90	-4
Rock music concert	100	6
Aircraft take-off and landing	125	31
Limit for damage to the hearing (CEC limit for long time exposure)	85	-9

6.7.5 Induced noise on subscriber lines

Power and traction lines can introduce hum of the fundamental power frequency and its harmonics on subscriber lines. (The mechanism depends partly on the degree of unbalance to earth in the subscriber circuit). According to an old rule-of-thumb, Recommendation G.123 [13] states that the psophometric e.m.f. (i.e. open voltage) of the line terminals, to which the telephone is to be connected, should not exceed 1 mV.

If one assumes, for the sake of simplicity, that the impedance of the set is 600 Ω and that the voltage is decreased by half when the set is connected, the induced noise limit becomes -64 dBmp at the terminals of the connected set.

Assume, as an example, that the line attenuation is 3 dB and that the receive and transmit pads of the local digital exchange are respectively R = 7 dB and T = 0 dB. Subscriber cables as a rule are very well balanced to earth and the same applies to normal telephone sets which have no direct connection to earth. The unbalance in a subscriber circuit is usually caused by the unbalance at the input port of the digital exchange and it is there that the signal conversion from longitudinal to transversal voltage takes place. Thus, the equivalent circuit noise at the 0 dBr points in the exchange are:

In the receive direction -64 + 3 + 7 = -54 dBm0p;

In the transmit direction -64 + 3 = -61 dBm0p.

6.7.6 The subjective effects of non-stationary noise

In general, frequently occurring intermittent noise is more disturbing than continuous noise of the same instantaneous level. Therefore, in systems where intervals of noise and silence are apt to be interleaved, one often introduces so-called "comfort noise" in the noise-free intervals. This methodology has been used in analogue mobile systems and in some echo canceller applications. It is also used in digital mobile systems (e.g. see subclause 7.4.6 of I-ETS 300 038 [130]).

The effects of impulsive noise is under renewed study in ITU-T Study group 12 (SG12).

For further details, see annex B.

6.8 Power handling and non-linear distortion

6.8.1 Amplitude characteristics of the speech signal

According to information from ITU-T SG 12, an "average talker" produces, at a 0 dBr point, a mean active speech level of

$$L = -11 - SLR \qquad \text{dBm0} \tag{6.8.1}$$

where *SLR* is referred to the 0 dBr point in question and the speech level is measured with an instrument designed in accordance with ITU-T Recommendation P.56 [112].

Moreover, taking account of the variation among talkers and their handset handling, the standard deviation of the mean active speech level at a 0 dBr point is about 5 dB.

The instantaneous speech signal is more "peaky" than white noise. According to D. L. Richards, "Telecommunication by Speech" [108], the statistical distribution of the (absolute) voltage V can be simulated by the function

$$P(X) = \frac{K}{\Gamma(M)} \cdot (KX)^{M-1} \cdot e^{-KX} \tag{6.8.2}$$

where

X = V/Veff; Veff = r.m.s. (mean) value
M = a constant, about 0,5 for modern, linear microphones.

$$K = \sqrt{M(1+M)} \tag{6.8.3}$$

Equ.(6.8.2) is to be interpreted as follows:

> The probability to find a value in the interval (X, X + dX) is P(X)dX.

The cumulative statistical distribution F(X) is obtained by integration of equ.(6.8.2). It is depicted in figure 30. For comparison, the distribution for the envelope of white noise is also shown.

NOTE: Figure 3 of ITU-T Recommendation P.50 [107], gives the instantaneous amplitude distribution of an artificial voice signal. Note, however, that this applies to that signal, not necessarily to real voice signals.

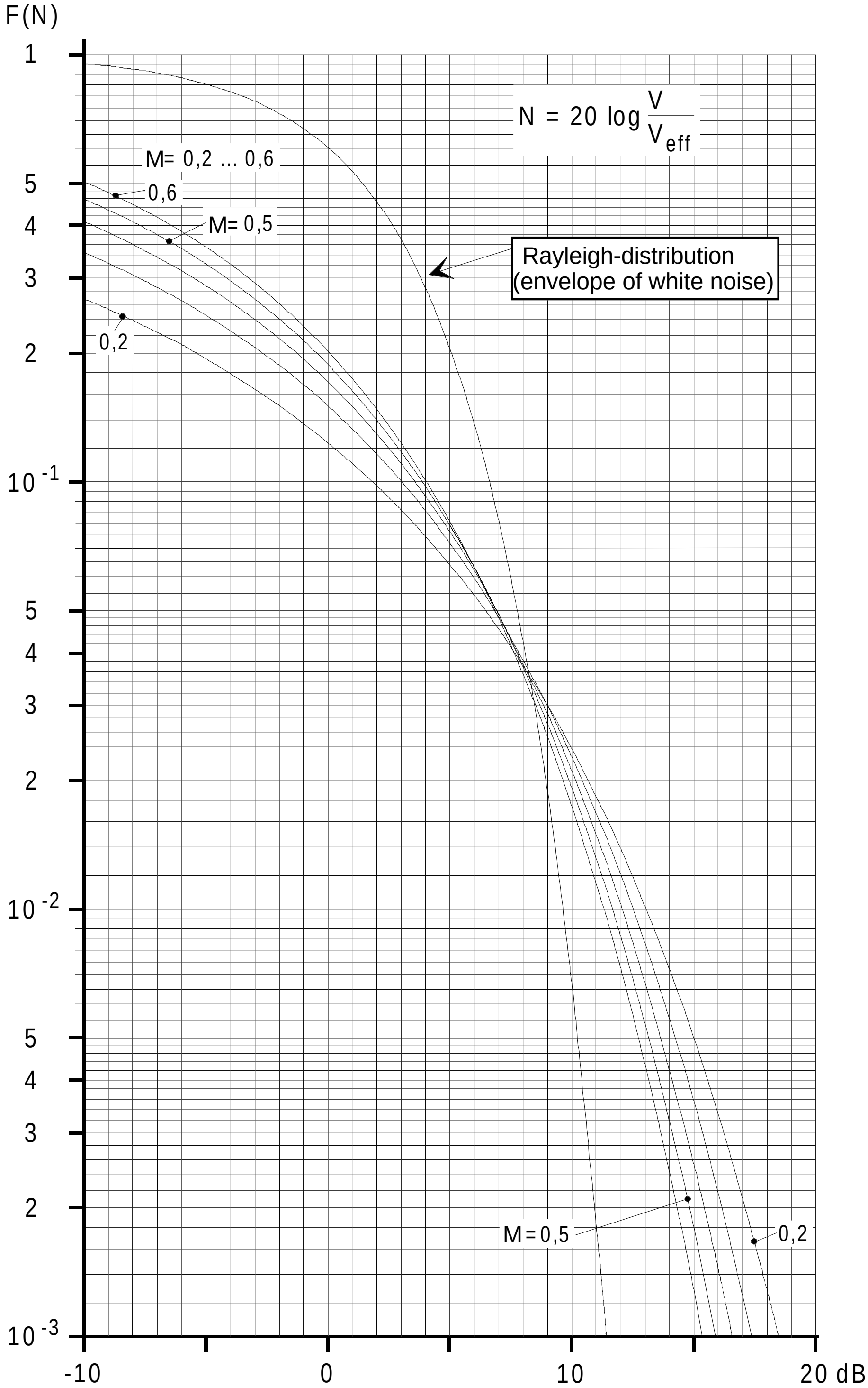


Figure 30: Cumulative statistical distribution of speech signals and of white noise envelope -Abscissa = N = 20lg(V/Veff)

Measurements have shown that peaks can occur which are 18 dB higher than the mean r.m.s. However,subjective tests have indicated that those peaks can be clipped 6 dB without noticeable quality degradation which corresponds to a clipping at a level 12 dB higher than the mean r.m.s. (This means clipping about 1 % of the time, which figure is obtained by integration of equ.(6.8.2)).

6.8.2 Power handling, FDM systems

FDM (Frequency Division Multiplex (carrier)) systems are sensitive to the *total* signal power loading.

Large-capacity FDM systems are designed to allow, in an up-modulated band, a long-term average power of -15 dBm0 per channel, taking into account signalling, carrier leaks and speech pauses.

The so-called activity factor is usually taken to be 25 %. Referring directly to actual speech during active periods, the corresponding mean active speech level becomes -11 dBm0.

This relation does not take account of the transmission of non-voice services. Many recommendations for such services are based on the assumption of a limit of -13 dBm0 on the 1-minute mean power.

Note, however, that FDM systems with fewer than 240 channels must be designed for a higher long-term average power per channel. Thus, a 12-channel FDM system should be able to handle -7.5 dBm0 per channel instead of -15 dBm0.

6.8.3 Power handling, PCM systems

For equipment using 64 kbit/s PCM, the power handling capacity is directly connected with the coders and decoders. The PCM maximum coding level corresponds to just clipping of a sinusoidal signal at +3,14 dBm0 for ideal codecs using the A-law and at +3,17 dBm0 for the µ-law. (The real peak signal levels are of course 3 dB higher).

What margins exist against noticeable peak clipping of actual speech signals in the network?

By using equ.(6.8.1) the margin C dB against "just noticeable" peak clipping (at 12 dB higher than the mean), and the corresponding percentage Pc of talkers subjected to it, can be calculated. Thus,

For the *nominal* SLR = 7 dB: C = 12 dB, Pc = 0,8 %
For the *minimum* SLR = 2 dB: C = 7 dB, Pc = 8 %

6.8.4 Measurements of actual signal levels in networks

For several reasons, the actual speech levels in networks is being studied anew in ITU-T. Since 1970 analogue transmission has to a large extent been replaced by digital transmission and there is a proliferation of approved telephone sets, many of them having DC feed sensitivity regulation. For information, some results from a recent investigation in the Swedish network are cited (COM 12-27: Measurements of speech levels; Telia, Dec. 1993. [126], see tables 5 - 9).

Measurements were made on outgoing calls from one local exchange for local, trunk and international calls. The exchange has a greater proportion than average of business subscribers. (Investigations will continue with measurements in many other points in the network).

Table 5: Speech levels

	Local calls	Trunk calls	International calls
Mean	-17,2 dBm0	-17,4 dBm0	-15,5 dBm0
Standard deviation	5,1 dB	5,1 dB	5,3 dB
Number of calls	1 475	1 117	1 061

Table 6: Levels of other signals (mostly fax and data signals)

	Local calls	Trunk calls	International calls
Mean	-13,7 dBm0	-13,3 dBm0	-14,0 dBm0
Standard deviation	3,1 dB	3,1 dB	2,7 dB
Number of calls	101	223	1 012

Table 7: Type of signal, % of all calls

	Local calls	Trunk calls	International calls
Speech	60	64	40
Signal (fax/data)	4	13	39
Unspecified	36	23	21

Table 8: Activity factor, mean % signal time to total time

	Local calls	Trunk calls	International calls
Speech	43	48	46
Signal (fax/data)	61	69	64

Table 9: Call duration, mean values

	Local calls	Trunk calls	International calls
Speech	3 minutes 44 s	4 minutes 51 s	6 minutes 20 s
Fax/data	2 minutes 53 s	2 minutes 8 s	1 minutes 47 s

6.8.5 Effects of non-linear distortion

Information on this topic is given in Annex F to the ITU-T Recommendation P.11, [26]. Figure 31 is a reproduction from this. As can be seen, cubic distortion is much more detrimental to the speech quality than the quadratic one. (As is well known, a carbon microphone has a predominantly quadratic distortion).

Note that these curves do not necessarily apply to non-linear distortion in a sidetone path.

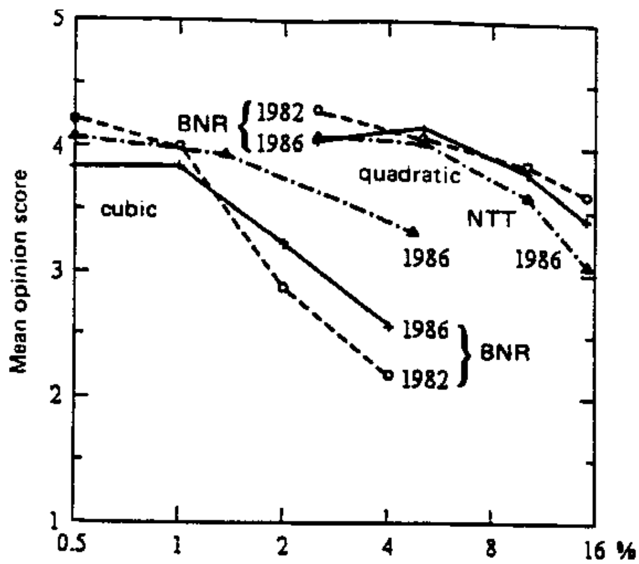


Figure 31: Subjective ratings for non-linear distortion

6.9 Crosstalk

6.9.1 General

In modern equipment, crosstalk should not be a problem. However, in some exceptional circumstances it might be necessary to check that the crosstalk attenuation is sufficiently high. Intelligible crosstalk should of course never be allowed and even just audible crosstalk can be rather annoying.

The subjective effects of crosstalk are treated in CCITT Recommendation P.16 [27] and possible crosstalk paths are discussed in CCITT Recommendation G.105 [12]. Annex A to CCITT Recommendation G.134 [15] describes some suitable methods for measuring crosstalk. CCITT Recommendation G.151 [16] gives limits for crosstalk.

*Between circuits***:** The circuit performance objective for the near-end or far-end crosstalk ratio (intelligible crosstalk only) measured at audio frequency at trunk exchanges between two complete circuits in terminal service position should not be less than 65 dB.

*Between the go and return channels of a 4-wire circuit***:** The circuit performance objective for the near-end crosstalk ratio between the two directions of transmission should be at least 43 dB. In certain cases, such as when speech concentrators and echo control devices are used, at least 58 dB and 55 dB respectively.

6.9.2 Subjective effects of crosstalk characterized by Crosstalk Receive Loudness Rating (XRLR)

Annex A to ITU-T Recommendation G.111 [34] gives the expression for the weighted crosstalk attenuation *Lx* in the speech band to

$$\mathrm{Lx} = 1{,}4 - 10.\,\mathrm{lg}\left\{ \int_{500}^{2\,000} 10^{-\mathrm{L}/10}.\frac{\mathrm{df}}{\mathrm{f}} \right\} \; dB \qquad (6.9.1)$$

where *L* is the frequency-dependent crosstalk loss.

In many cases, *Lx* is very nearly equal to the loss at 1 020 Hz.

In CCITT Recommendation P.16 [27] the crosstalk phenomena are referred directly to the terminals of the disturbed telephone set. Here, in order to simplify the matter, we will instead make the (nearest) 0 dBr point the reference for crosstalk attenuation, circuit noise, equivalent circuit noise and loudness ratings.

Suppose there is a crosstalk path with the (weighted average) crosstalk attenuation *Lx* between a sending point with the relative level *Ns* dBr and a receiving point with the relative level *Nr*. At the 0 dBr point, the equivalent crosstalk attenuation is

$$Lxo = Lx + Nr - Ns \qquad (6.9.2)$$

In the case of many possible crosstalk paths there is only a need to consider the worst one.

At the 0 dBr point, the crosstalk receive loudness rating is

$$XRLR = Lx + RLR \qquad (6.9.3)$$

where RLR also is referred to the 0 dBr point.

Crosstalk is partly masked by circuit noise and the equivalent circuit noise resulting from room noise at both send and receive side. The room noise components can be computed by the equations given in the description of the ETSI model. As a typical example, consider the following case:

The disturbed channel has at the receive side RLR = 3 dB, LSTR = 10 dB, room noise = 40 dB(A). The send side has SLR = 7 dB, Ds = 3 and room noise 40 dB(A). (RLR, SLR are referred to the 0 dBr point, Ds is the weighted average of the send sensitivity between direct and diffuse sound). The equivalent circuit noise caused by room noise becomes

From receive side:	Nor	= -74 dBm0p;
from send side:	Nos	= -69 dBm0p;
total:	No	= -68 dBm0p.

However, for reasons of safety margins, only the real, electric circuit noise should be considered in crosstalk evaluations.

The subjective effect of crosstalk is of course dependent on the speech level of the disturbing talker. In a 0 dBr point the transmission planning ensures that the speech levels should lie within certain limits. Measurements on a number of circuits indicate that the *mean* active speech level is about -18 dBm0, with a standard deviation of 5 dB.

Figure 32 (reproduced from CCITT Recommendation P.16 [27]) gives the required crosstalk receive loudness rating XRLR as function of the circuit noise Nc, for the median curves of "audibility" and "intelligibility" thresholds. The standard deviation is in the order of 5 dB also for the listening sensitivities.

The total standard variation is thus in the order of 7 dB. By means of the curves in figure 32 it is possible to estimate the risk for audible and intelligible crosstalk. For instance, for a circuit noise of Nc = –65 dBm0p, RLR = 3 dB and a risk of intelligible crosstalk of less than 0,1 % (= 3xSt.D), the required crosstalk attenuation becomes *Lxo* > [53 - 3 + (3x7)] = 71 dB.

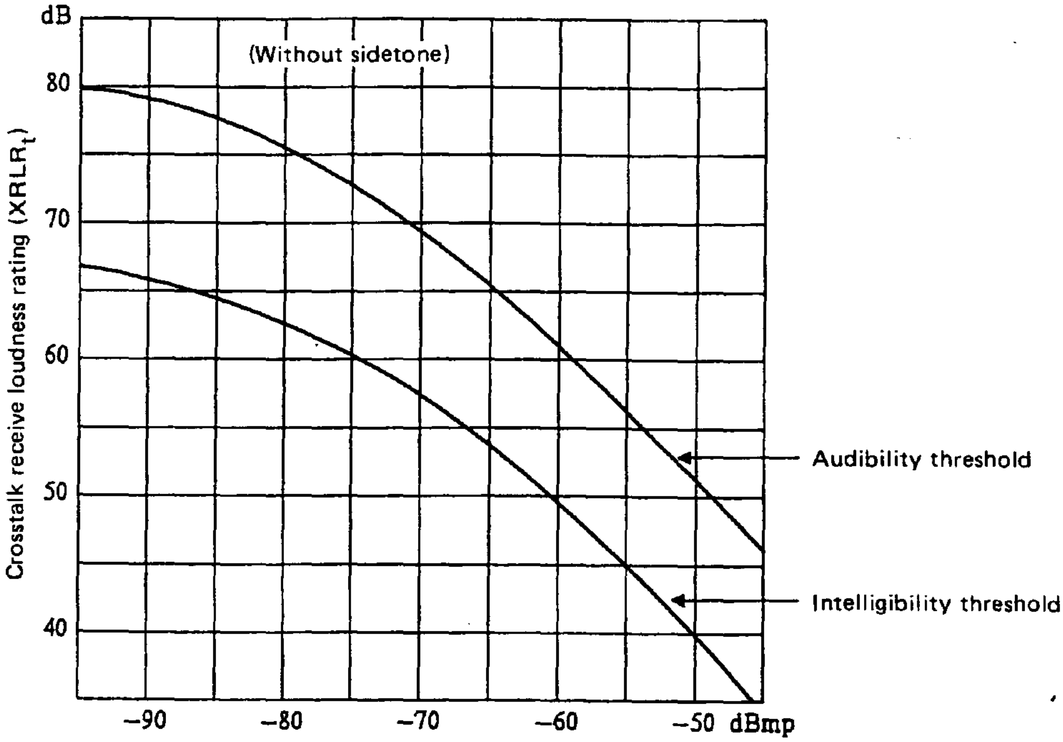


Figure 32: Threshold values of crosstalk receive loudness rating XRLR (referred to a 0 dBr point)

6.10 Effects of time-varying impairments

It is very difficult to incorporate the effects of time-varying parameters at this stage of the development of the model. Time-varying parameters are only recently being studied. Some examples of the origin of time-varying parameters are:

- new signal processing techniques in terminals (adaptation processes in echo-cancellers);
- the dependence of the relative position of the receiver in a radio path yielding differences in field strengths and interferences.

At the moment these influences are being studied at different laboratories and no subjective results are available yet.

7 Equipment and systems with special transmission properties

7.1 Echo control devices

This subclause describes only the principle of Echo Suppressors (ES)s and Echo Cancellers (EC)s presently in use. Additional information is given in annex D ("Technical report on echo cancelling"), and in CCITT Recommendation G.164 [18] for ESs and Recommendation G.165 [19] for ECs.

Presently, echo control devices are used for suppressing echo in connections with a mean one-way transmission time of more than 25 ms, i.e. mainly in international connections. The main task of an echo control device (ECD) is to suppress or compensate signals coupled within the echo path, without inserting remarkable impairments to the speech quality. Two main basic principles are known and in use since several years, the echo suppressor and the more modern one, the echo canceller.

7.1.1 Echo suppressors

As shown in figure 33, an ES is able to suppress the sending path (S_{in} to S_{out}) and to insert a loss in the receive path (R_{in} to R_{out}). The signal amplitude is detected from the send and receive path and compared in a logic circuit and used to control the loss in both directions. If a signal from party A above a defined threshold level is present in the receive path at Rin, the ES changes to suppression mode, i.e. a high loss of about 50 dB is inserted in the send path, suppressing every possible coupled signal.

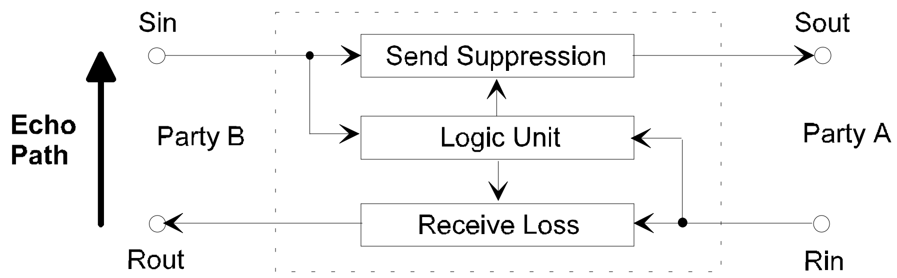


Figure 33: Echo suppressor

This mode of course is encountered by party A subjectively as an interruption and party B would be unable to 'break-in' during a conversation. For those situations, also called 'double talking', a differential circuit is used to compare the speech level in send and receive path. If double talking is detected, the send suppression is removed and a loss in the range of 5 dB to 15 dB is inserted in the receive path. Several other technologies such as fixed or adaptive differential sensitivity are used to operate also under 'break-in' conditions. Echo suppressors are applicable for echo path delays up to about 25 ms.

7.1.2 Echo cancellers

The principle of an echo canceller (EC) is shown in figure 34. A received signal from party A is modified by the echo estimator, synthesizing a replica of the echo path and subtracting this signal from the send path. ECs are available for analogue 4-wire and digital paths. They can also be characterized whether the signal processing and subtraction is by analogue or digital means.

Since the echo path varies for every connection mainly in delay and distortions, the process of converging to the new echo path must be fairly rapid, e.g. well below 1 second. ECs are more advantageous than ESs. Since no suppression in the send path or insertion of loss in the receive path is used, the receive direction of the talker does not seem to be interrupted for single talk conditions in contrary to an ES. During break-in and double talk conditions, basically echo cancellation is continued, but the echo estimator attempts to adopt to this 'new echo signal' and may cause a degradation of speech quality and reduction of cancellation. However, several algorithm are used to avoid these effects.

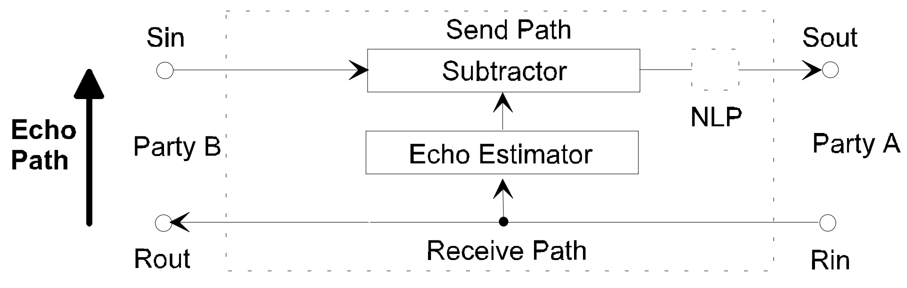


Figure 34: Echo canceller

At the output of the send path ECs may be equipped with an additional unit called "non-linear processor" (NLP) or "center clipper". The aim of this device is to provide a suppression of residual echo levels below a defined threshold. It must be noted, that the operation of the NLP may cause an annoying effect due to environmental noise at party B. If party A is talking, the NLP is operating causing a "silent" receive. If party A is listening and party B is talking, then the NLP is inoperative and the environmental noise is transmitted together with the speech signal.

To avoid those annoying intervals of speech with noise followed by an interval of silence, some artificial noise called "comfort noise" can be inserted, or some of the background noise is allowed to pass through the NLP.

7.1.3 Planning rules for ECDs

Common to all ECDs is the requirement for a minimum echo path loss of 6 dB, corrected to equal relative levels, for proper operation. In situations with lower values than 6 dB, the coupled signal will be considered as near-end speech and the adaptation process is stopped.

The problem for a correct "double talk detection" also requires a careful planning of speech levels in a network. In mixed analogue/digital networks situations may exist, where the hybrid forming the echo path is terminated in a long analogue 2-wire section with the respective high loss (party B in figure 34). If the speech signal at R_{in} is high, due to a fully digital routing to party A, the coupled echo signal at S_{in} could be higher than the near end speech signal from party B and it becomes difficult to distinguish between single and double talk conditions. Furthermore an EC tries to adapt to the lower signal resulting in a degradation of the cancelling process.

Careful network planning is also required with respect to very high speech levels, to avoid nonlinearities arising from overloading. This could cause problems for some ECs in their adaptation process, double talk detection and control of the NLP.

Echo cancellers are designed assuming a linear echo path. Nonlinearities in the echo path like voice companding systems and codecs with bit rates below 64 kbit/s may affect the performance of the EC. In future fully digital connections the electrical echo path as provided by a hybrid, will be replaced by an acoustical coupling as in digital telephone sets, especially in the handsfree mode.

Private networks may exceed the established limits for transmission time in the terminal equipment area, same as for specific coding principles as used in cordless or mobile applications.

Regarding these facts, it seems to be impossible that existing ECDs can provide a sufficient echo performance in all cases. Therefore separate echo cancellation for these specific applications must be taken into account. This will lead to an increasing use of ECDs in tandem, but practical experiences have shown that ECs designed according to CCITT Recommendations G.165 [19] will provide no major cause for complaint. As far as ESs and ECs are designed according CCITT Recommendations G.164 [18] and G.165 [19], compatibility is also guaranteed between an ES inserted at the one end and an EC at the other end.

Nevertheless, the tandeming with specific echo control devices (e.g. acoustic echo control, soft suppressors etc.) needs further study.

7.2 Low bit-rate codecs

7.2.1 Introduction

In many applications, such as, for example, mobile communications and voice storage services, the available channel capacity is limited, or it is found economical to use the least number of bits for coding speech.

Moreover, the imminent explosion of many new types of services, e.g. mobile and multimedia, interactive and non-interactive, using single-point and multi-point connections, will imply the introduction of newly-developed types of speech coding algorithms, operating at very low bit-rates.

New transmission technologies will soon consider, in addition to "narrow" telephone bandwidth (3,1 kHz), "wideband" digital coding of telephone speech (50 Hz up to 7 kHz). Also "audio" coding technologies (20 Hz to 20 kHz) are likely to be specified in ISO (MPEG) and ITU-T within a few years time.

Activities have been announced within standard bodies like ISO (MPEG) or ITU-T (SG 12 Speech Quality Experts Group and SG 15) in order to consider the development and the performance of some new standard coding algorithms capable of encoding up to audio signals of different bandwidth at different bit rates in actual network conditions (e.g. including error probabilities). Speech quality categories often used to describe codec performance are "toll quality", to indicate the average quality of a long-distance PSTN connection, "communications quality", characterised by good intelligibility, speaker identity maintained, but noticeable loss in quality, and "synthetic quality", where the speaker identity is mostly lost, and the speech sounds synthetic in nature (see figure 35).

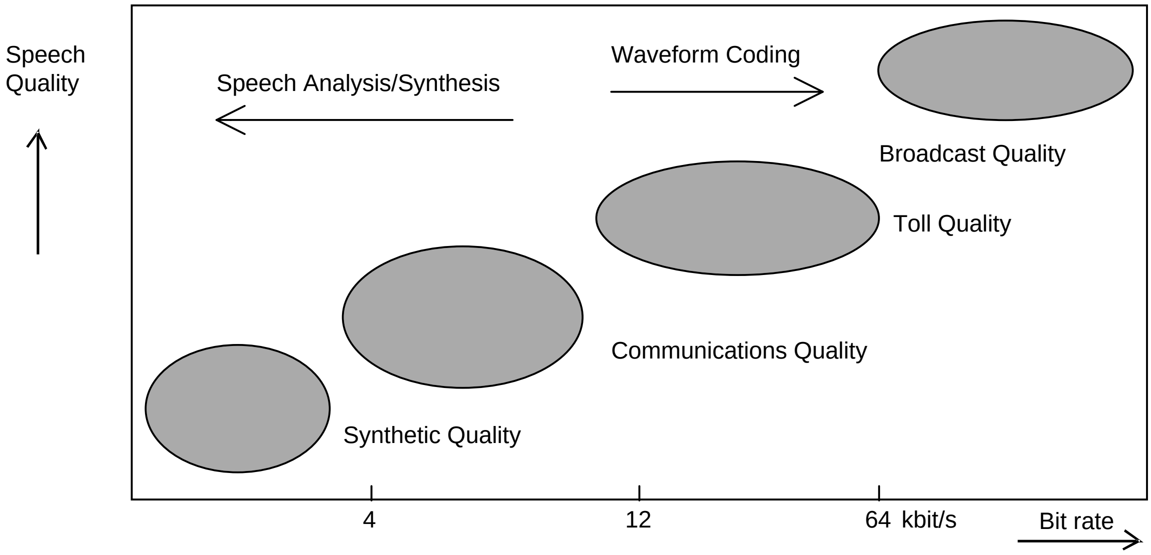


Figure 35: Quality of speech coding transmission rates

The codec performance (see also subclause 9.2.7 and annexes E and G) could be expressed in quantities as:

a) equivalent quantization distortion units or qdus (by definition, 1 qdu = 1 "commercial" PCM G.711, 2 qdus = 2 PCM G.711 in tandem, etc. where G.711 refers to CCITT Recommendation [62]); or

b) equivalent Q (dB) values (multiplicative S/N); or

c) Mean Opinion Scores (rating scale from 1 to 5), or percentages of Good or Better (GOB) and Poor or Worse (POW), taking into account the "context" of measurement.

The ETSI behavioural model could play a new role in such topic, by expressing the degradation introduced by the insertion of low bit-rate codecs in the network in terms of a well quantified "equipment impairment factor", as already proposed and accepted in ITU-T SG 12.

For this purpose it is essential to consider all pieces of available information concerning objective and subjective evaluations of the performance of any newly marketed terminal or service utilising speech encoding techniques.

7.2.2 ITU-T PCM, ADPCM and LRE (Low Rate Encoding) algorithms

The following speech coding algorithms have been recommended by ITU-T (former CCITT):

- CCITT Recommendation G.711 [62]: this 64 kbit/s Pulse Code Modulation (PCM) of voice frequencies was published in 1972 and furtherly amended. It has two well known encoding laws, that are commonly referred to as the A-law and the µ-law;

- CCITT Recommendation G.721 [63]: this 32 kbit/s algorithm is a further development of the CCITT Red Book version, and has a 15 level quantiser, so that the ‘0000’ code cannot be transmitted;

- CCITT Recommendation G.723 [65]: this recommendation contains two fixed rate algorithms for use in DCME (24 kbit/s for the encoding of overloaded channels normally working at 32 kbit/s, and 40 kbit/s algorithm for better quality of voice band data);

- ITU-T Recommendation G.726 [121]: this recommendation describes a family of ADPCM algorithms at 16, 24, 32, and 40 kbit/s fixed rates. In addition to the use in DCMEs, the 32 kbit/s algorithm has been chosen for use in DECT (Digital European Cordless Telephone) and CT2 (Cordless Telephone of 2nd generation). This recommendation replaces CCITT Recommendations G.721 [63] and G.723 [65];

- ITU-T Recommendation G.727 [66]: this recommendation adds to ITU-T Recommendation G.726 [121] a technique known as 'embedded coding', which allows the least significant bits to be stolen periodically without affecting severely the voice quality by the decreased number of bits transmitted. It is useful in packetised systems to permit congestion control;

- CCITT Recommendation G.722 [64]: this recommendation is based on ADPCM algorithms, applied to two bands encoded separately, for a 7 kHz wide band transmission at 48, 56 and 64 kbit/s. It is used for audioconferencing, broadcasting commentary channels and ISDN telephony;

- ITU-T Recommendation G.728 [81]: this recommendation contains a 16 kbit/s algorithm using a Low-Delay (less than 2 ms) Codebook Excited Linear Prediction (LD-CELP), with the same quality of ITU-T Recommendation G.726 [121] at 32 kbit/s. Main foreseen application is for audio-visual services.

An additional feature of many ADPCM algorithms is the synchronous tandeming capability, which permits the minimising of the accumulation of distortion in multiple stages of conversion between PCM (using the same encoding law and in absence of errors) and ADPCM, and vice versa.

- ITU-T Recommendation G.763 [82]: DCME (Digital Circuit Multiplication Equipment) using 32 kbit/s ADPCM and Digital Speech Interpolation;

- ITU-T Recommendations G.764 [83] & G.765 [84]: Voice Packetization - Packetized Voice Protocols & PCME (Packet Circuit Multiplication Equipment).

Other ITU-T algorithms under study are:

- ITU-T Recommendation G.728 [81], at variable bit-rate for DCME applications, robust to errors (random and burst) for FPLMTS and PCME;

- 8 kbit/s and 4 kbit/s speech codecs for personal communications;

- wide band coding up to 16 kbit/s (7 kHz) without annoying effects for music signals, and audio coding (15 Hz to 20 kHz);

- finally, an ITU-T algorithm, for audio-visual telephony, for coding speech at very low bit-rate (around 6,4 - 6,8 kbit/s) is likely to be recommended.

These are all algorithms under study, in collaboration between SG 12 and SG 15, in both the ITU-T and ITU-R sectors.

7.2.3 ETSI speech coding algorithms

GSM 900 (Global System for Mobile communication) is a public digital mobile radio system that works in the range 890 MHz - 915 MHz for the mobile to base station direction, while another range (935 MHz - 960 MHz) are reserved for the base station to mobile direction. Its evolution is the DCS 1800 (Digital Communication System), which allows the use of more radio channels with a substantially unchanged protocols.

Transmission aspects of GSM 900 are reported in GSM 5 Series of specifications which are also published as I-ETSs (Interim European Telecommunication Standards).

Speech processing aspects (bit-rate at 13 kbit/s) and related transmission properties are contained in GSM 6 Series, (also I-ETSs):

- I-ETS 300 036 [128]: "European digital cellular telecommunications system (phase 1); Full-rate speech transcoding (GSM 06.10)";

- I-ETS 300 037 [129]: "European digital cellular telecommunications system (phase 1); Substitution and muting of lost frames for full-rate speech traffic channels (GSM 06.11)";

- I-ETS 300 038 [130]: "European digital cellular telecommunications system (phase 1); Comfort noise aspects for full-rate speech traffic channels (GSM 06.12)";

- I-ETS 300 039 [131]: "European digital cellular telecommunications system (phase 1); Discontinuous transmission (DTX) for full-rate speech traffic channels (GSM 06.31)";

- I-ETS 300 040 [132]: "European digital cellular telecommunications system (phase 1); Voice activity detection (GSM 06.32)".

All standards related to the Half-rate GSM speech codec for traffic channels have been published in ETS 300 581 [127]. The half rate codec performs as well or slightly worse than the full rate codec in most conditions. The exceptions are for background noise and tandem configurations, where the performance of the half rate GSM codec is considerably worse than that of the full rate GSM codec.

Other European future systems that will imply the possible use of speech coding algorithms at a reduced bit-rate are the TFTS (Terrestrial Flight Telephone System), the UMTS (Universal Mobile Telecommunications System), and the TETRA (Trans European Trunked RAdio), a private mobile radio network for independent groups of users, that may be connected both to other TETRA groups or to the PSTN, PDN, ISDN and PBX.

Since 1992, the DECT (Digital European Cordless Telecommunications) has been adopted as the European cordless standard, and described in the ETS 300 175 [47], and I-ETS 300 176 [133]. DECT uses for speech coding the ITU-T algorithm ADPCM at 32 kbit/s, and the bandwidth allocated is of 20 MHz (1880 - 1900 MHz); the perceived quality from the user is affected by poor C/I (Carrier-to-Interferer) power ratios and by discontinuities during the handover phase between DECT islands; therefore, as in DCME and PCME systems, the transmission properties that should be considered, ought to take into account also the interactions between the algorithms utilised by the system and special features of the system itself. Also the interactions of radio channel coding with source speech coding in mobile systems have a relevant influence on the overall quality, as perceived by the users.

7.2.4 Other regional speech coder algorithms

It should be pointed out that other regional bodies outside of Europe (ETSI), but in particular in USA (CTIA) and Japan (JDC), have on-going activities for the deployment of new algorithms in the wireless, satellite, or other communications fields at low bit-rate in public or private networks. These algorithms are not considered explicitly in this section, for sake of brevity, but they do (or will soon) exist, and have an impact on the proliferation of systems that interwork and influence the overall quality of the world-wide distributed networks.

Among the most interesting future possible developments there are:

i) the DOD (US Department Of Defence) possible evolution of present military standard LPC10 and commercial LPC++, with synthetic speech at 2,4 kbit/s towards a target at the same bit rate but with the quality (synthetic/communications) of the present CELP US DOD at 4,8 kbit/s (US federal standard secure voice proposed standard FS 1016 [156]);

ii) the evolution from Inmarsat-M land mobile satellite codec IMBE (Improved Multi Band Excitation) working at 4,15 kbit/s to a new standard called MINI-M;

iii) the evolution of the proposed standard IS 54 [69] (US 8 kbit/s cellular standard VSELP (Vector Sum Excited Linear Prediction) speech coding algorithm) towards half rate digital cellular telephony standards.

7.2.5 General: New low bit-rate codecs and their properties

The success of the studies on digital systems for the representation of the speech signal pushed the research towards lower and lower bit-rates, by utilising the models developed for the mechanism of production and perception of natural speech.

In parallel, techniques of analysis and elaboration of signals open new perspectives for the control of noise and other sources of degradation in telecommunication networks, that may be taken into account in the codec developments.

The main impairments which may occur in the encoding process are:

- increased quantisation distortion, either in a fixed or variable amount, depending on both the signal and/or the bit-rate changes (in variable bit-rate codecs);
- bandwidth restriction: for example, this is a feature of one ADPCM algorithm designed specifically for voice band data transmission;
- speech level dependency effects: both input and listening level ranges are important;
- talker gender dependency effects: whether the talker is male, female or a child, for example, the speech may be degraded differently;
- background and environmental noise effects including background music, that are more and more taken into consideration for personal communication systems;
- error pattern and random/burst error distribution effects;
- interworking with other algorithms and number of transcodings;
- crosstalk (or multiple talker) effects;
- delay (and related echo effects);
- range of listening levels.

Speech coding algorithms can be classified into the following types (see figure 36):

- waveform coding, the aim being to reproduce the original waveform as accurately as possible, that has acoustic robustness but rather high bit-rates;
- vocoders (<u>vo</u>ice <u>coders</u>) for very low bit-rates, or parametric coders, that use a speech production model to transmit perceptually important parameters only;
- hybrid coding, that combines features from both waveform coders and vocoders, to get efficient speech coding and good quality.

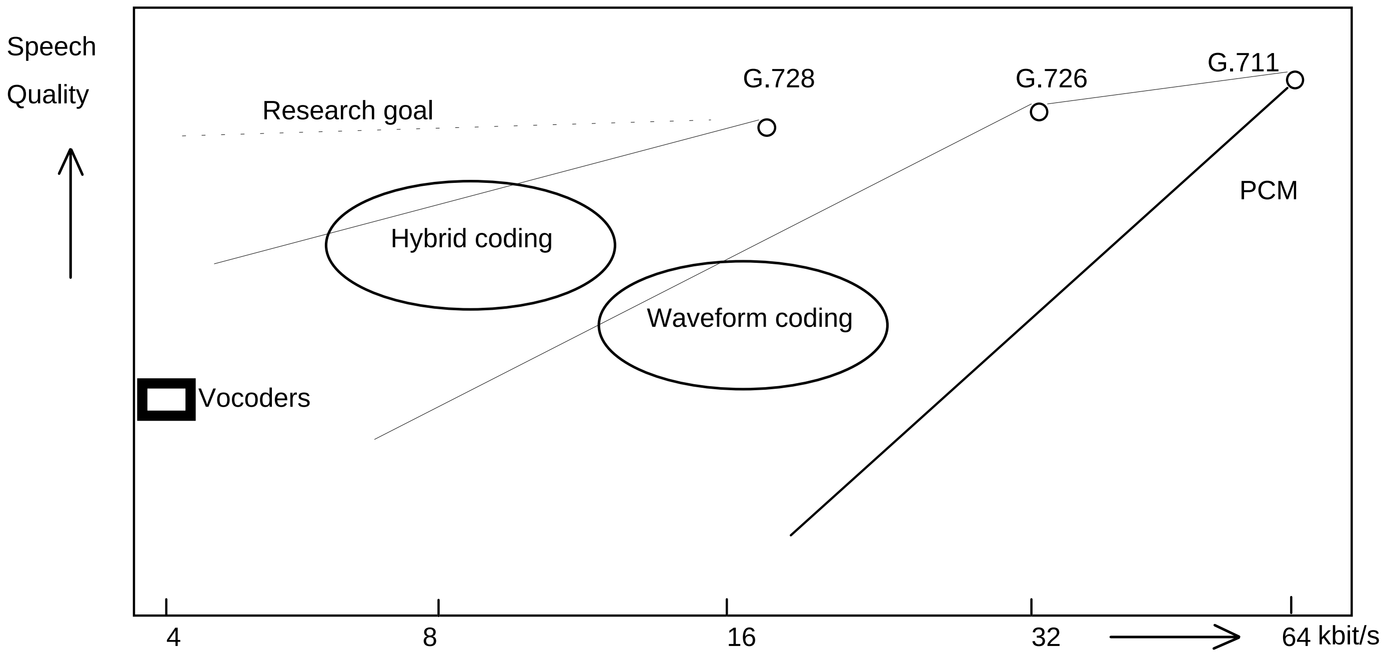


Figure 36: Quality of telephone speech as a function of the bit-rate

As waveform coding (even in the frequency domain, e.g. SubBand Coding (SBC), or Adaptive Transform Coding (ATC), or Adaptive Predictive Coding (APC) is no longer used for low bit-rate algorithms, 16 kbit/s downwards, the difference between the newly proposed schemes and PCM is not just expressed by the increased amount of quantizing noise; therefore, new types of distortions ought to be considered, e.g. longer delays and lack of naturalness, or transmission error effects, caused by bursts or missing cells, that may cause severe degradation of the ‘perceived quality'.

The key issue in designing codecs at low bit-rates, where only a few bits per sample are available, is finding bit-efficient representations of the typical parameters of speech production models (see figure 37).

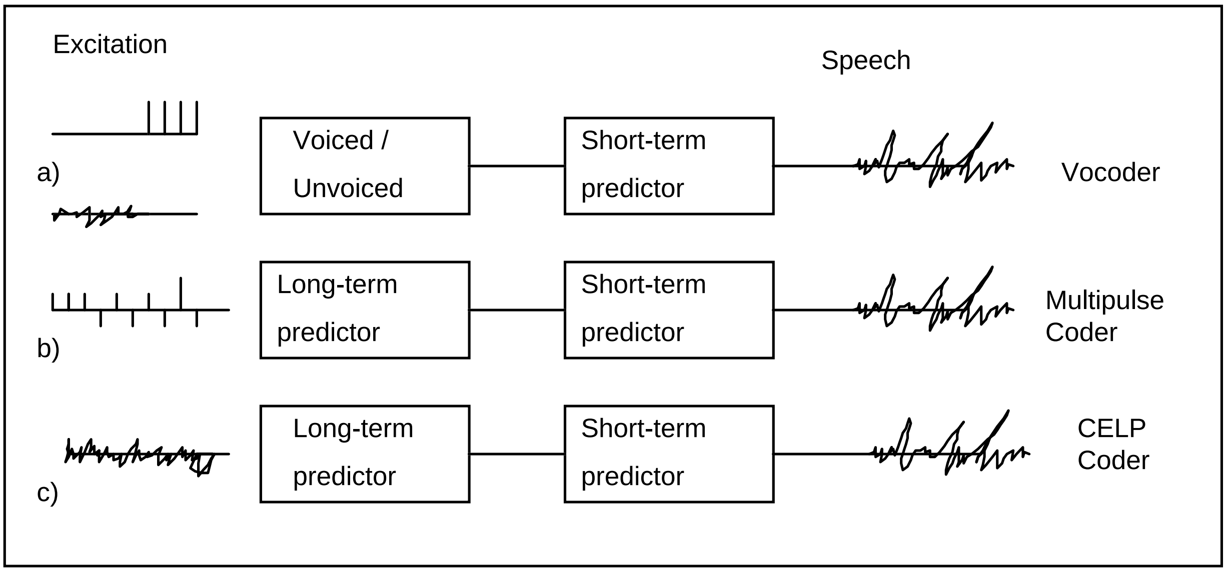


Figure 37: Models of speech exitations in vocoders (a) and hybrid coders (b, c)

This is obtained by comparing a completely synthesized speech segment with the original one, i.e. by computing the error signal that produces, for a given synthesis filter, an output perceptually close to the original source signal. The choice of the most appropriate excitation signal is made by considering both the reconstructed speech over several samples, and the weighted difference from the original speech that is obtained by minimising suitable distortion criteria; this procedure is called **analysis-by-synthesis adaptive predictive coding**.

Two low rate encoding algorithms based on the LPC model are the Multi-Pulse Excited (MPE, where the residual signal is quantized by a sequence of pulses), and the Code Excited Linear Predictive (CELP) coding schemes, the latter often referred to as the stochastic coder, that represents the excitation signal by a random sequence of numbers, where the best sequence is chosen from a codebook of stored sequences using an iterative procedure. Nowadays these coding techniques are applied in the range from less than 4 kbit/s to 16 kbit/s, and should present a performance rather independent of voice characteristics, languages, sex of talkers, input levels within a given range, etc.; these algorithms should also be robust to multiple voices and background noise and should work well with poor signal to noise ratios; to this aim perceptual properties and hearing mechanisms are exploited by codec designers.

Many of the time-domain coding algorithms are used for telephony signals, while frequency-domain coders are better suited for wide-bandwidth audio signals (e.g. sub-band, harmonic coding, adaptive transform and multiband excitation coding).

7.2.6 Effect of coder parameters on the quality

The effect of the various coder parameters on the quality of the reconstructed speech signal is related to the trade-off between the number of parameters and the rate at which they are adjusted. In fact, the whole mechanism related to the perception of degraded speech is not yet completely clear; as an example, frequency masking experiments showed that the hearing system has only a limited capability to detect noise in the frequency bands in which the speech signal has high energy (formant regions). Besides, there is no magic formula for an appropriate error evaluation criterion, for which a better understanding of the talking and hearing mechanism seems to be needed. In the following a brief description of the most common codec parameters and their impact on the overall speech quality is given.

Short-term predictor filter parameters:

The short-term predictor can be considered as a model of the vocal tract; it is important to obtain a so called log-spectral distortion below 1 dB (rather inaudible) with reduced peak values (the most perceptible).

The predictor coefficients' choice influences also the codec sensitivity to transmission errors. It is also worth considering that high level noises may upset the predictors, as well as aliasing distortion introduced by audio interface circuitry.

Moreover, as coefficients are updated (every 10 to 20 ms), undesired transients may be introduced in the reconstructed speech.

Long-term predictor filter parameters:

The long-term predictor can be considered as a model of the periodicity in the excitation signal. The use of a long-term prediction filter improves the speech quality, specially for high-pitched voices; however, if it requires reducing the update rate or using differential encoding to reduce the number of bits, it may introduce peculiar and undesired effects.

Excitation parameters:

The excitation parameters are determined such that the mean squared error over a frame of samples is minimised. The excitation signal becomes more perceptually important as less and less is the bit-rate.

Post filtering:

Special techniques, such as post filtering, are sometimes used to try to reduce the loudness of the coding noise.

In principle, post filtering emphasises the perceptually more important spectral peaks in the speech signal, increasing the subjective signal-to-noise ratio, at the expense of signal distortion. If it includes a long term section, it can also mask the transient effects introduced by the (sudden) change of filter parameters. A disadvantage of post filtering is that it gives a reproduced distorted signal and this effect causes an even more noticeable degradation in performance of tandemed codecs. Recent tests have shown that substantial improvement in quality has not always been achieved.

Bit-rate:

These coders are used in a variety of applications in the range from 4 kbit/s to 16 kbit/s.

For communication applications it is generally desirable to maintain a constant bit-rate and the encoding is done frame by frame, while keeping the total number of bits per frame constant.

For some other applications, a variable bit-rate is adopted.

Quality is often considered as a monotone function of the bit-rate; the ITU-T recommendations contradict such view, as ITU-T Recommendation G.728 [81] at 16 kbit/s is not significantly worse than ITU-T Recommendation G.726 [121] at 32 kbit/s, and it can be stated the same for ITU-T Recommendation G.726 [121] at 32 kbit/s and ITU-T Recommendation G.711 [62] at 64 kbit/s.

Total delay:

The total delay, due to the encoding-decoding process, is a function of algorithmic delay (including error correction techniques), processing time and transmission time. In telephone conversations or real-time applications, a total delay of a few tenths of a millisecond can cause noticeable echo and objections by the customers.

Robustness:

Robustness is defined as the ability to cope with unexpected conditions. Possible examples are speech corrupted by background noise (tanks, helicopters and fighters represent limit conditions), and channel errors above 10^{-3} during transmission.

Application dependency:

In some applications, e.g. in portable communications systems, cost and power consumption are relevant parameters and could affect the choice of speech coding algorithms or coder parameters; this could influence indirectly the overall performance of the network, if other than quality factor is given a priority, e.g. in case of services delivered taking into account the quality/price ratio.

Non speech signals:

The performance of a codec is sometimes evaluated for non speech signals, like voice band data and DTMF tones or signalling tones. The optimisation of the codec for these requirements may affect the speech performance.

7.2.7 Objective and subjective speech evaluation measures

The simplest of the objective measures is the long term signal-to-quantization noise ratio (SNR), i.e. the ratio between the long term signal energy and the long term noise energy, the noise being defined as the difference between the input and the output samples. This measure is influenced by the high energy components of the speech signal.

An improved measure is the segmental SNR, where a logarithmic weighting converts SNR values into dB before averaging signal segments.

In the frequency domain, the logarithmic r.m.s. Spectral Distance (SD) measures the difference between the original and the reconstructed signal.

Anyway, all these measures tend to be inadequate for evaluating the performance of low bit-rate parametric speech codecs, due to the poor correlation with subjective judgements.

However, in ITU-T SG12 efforts are made to develop and recommend objective measures, based on human speech perception, that have a better correlation with subjective judgements.

The proposed recommendation, after passing the necessary validation test, will include perceptual speech quality concepts, expert pattern recognition techniques and well known measures in the time and frequency domain.

Formal subjective tests are described in ITU-T Recommendations P.80 [85] and P.83 [86].

7.2.8 Conclusion

The main features that new coding techniques use to exploit the redundancy of speech signal and the capabilities of speech production models for reducing the bit-rate have been highlighted. The impact of the choice of most parameters, including robustness, can be judged only by listening to the reconstructed speech.

In existing analysis-by-synthesis standard coders the speech quality varies, in terms of Mean Opinion Scores (MOS), using the Absolute Category Rating method with a 5-point scale, from about 3 (Fair) to about 4,5 (Good to Excellent).

Since the MOS are context-dependent, another method of evaluation of the codec performance has been internationally adopted, i.e. based on the equivalent speech to speech-correlated noise Q (dB), produced by a Modulated Noise Reference Unit (MNRU), which is transformable into quantization distortion units, or qdus, a simple concept very useful for network planners, as qdus possess the additivity property.

However, recent subjective test results showed the need of a revised "Quantisation Distortion Additivity Law" for analysis-by-synthesis coders; a number of proposals for new reference systems that sound more likely to new low bit-rate coders than the MNRU have been presented in ITU-T for consideration (T-Reference by Bellcore, S-Reference by NTT, PMNRU by Ellemtel).

Nevertheless it is suggested to consider the available data in the development of the ETSI model, in order to take into account the degradation of the overall quality of the network introduced by these new digital devices.

The actual proposal is to agree on an impairment figure to be assigned (provisionally), for each standard, in each working condition, to the term *Ie* contained in the proposed ETSI model (*Ie* is the impairment factor related to the inclusion of a certain equipment in the connection).

It is expected that ITU-T will produce a new recommendation on the matter of reference systems, that the ETSI model could then consider.

In general an increase in speech quality or decreases in coder bit-rate results in increasing the complexity of the coding scheme. Although no universally accepted formula based on ad hoc combinations of MIPS/MOPS figures and memory requirement has been established so far, how much power the implementation consumes is an important concern of coder developers that is being considered more and more (with speech quality and delay) for the comparison of different speech coding techniques. The increasing available processing power from digital signal processing devices has been the main driving force behind the development of new speech coding algorithms, together with the very high demand for mobile communications, that require good quality speech transmission (also in the presence of acoustic background noise and transmission errors), with the delay and the complexity kept within maximum limits.

Further reading: More detail related to subclause 7.2 can be found in references [1], [3], [48], [49], [53], [54], [55], [58], [59], [60], [61], [68], [74], [75], [79], [92], [94], [95], [101] and [102] (see clause 2, references of this ETR for more information on these references).

7.3 Digital Circuit Multiplication Equipment (DCME)

NOTE: Digital Circuit Multiplication Equipment (DCME) is described in ITU-T Recommendation G.763 [82].

7.3.1 General

Digital Circuit Multiplication Equipment, DCME, is used to enhance the capacity of digital transmission systems over links where a highly efficient traffic capability is required, e.g. a long distance link. A DCME has all of the following attributes:

- digital speech interpolation;
- low bit-rate encoding;
- dynamic load control arrangement in association with interfacing to a switching centre;
- capability to accomodate the following connection types:
  a) speech;
  b) 3,1 kHz audio (voiceband data and speech);
  c) 64 kbit/s unrestricted (transparent).

ITU-T Recommendation G.763 [82] deals with DCME using ADPCM (40, 32, 24 or 16 kbit/s) for the coding process. (The reduction, compression, of needed channels is in the order of 5:1, i.e. the circuit multiplication is 5).

A new Recommendation for DCME, using16 kbit/s LD-CELP for speech, is in preparation.

There are also on the market some multiplexing equipments for capacity enhancement which use proprietary algorithms for low bit-rate codecs. However, no performance statements can be given for such systems, regarding for instance delay times and distortions.

Annex H provides a tutorial overview of the transmission quality of a DCME which uses ADPCM.

In what follows some data will be given for a DCME that follows ITU-T Recommendation G.763 [82] (for more details, see this Recommendation).

7.3.2 Some data for a G.763 type DCME

Speech signals on telecommunication links are in general caused by two-way conversations between two talkers. It is customary for one talker to pause while the other speaks; thus, an active speech signal is present in each direction for the trunk channel during a fraction only of the available time. In addition, even when only one talker is speaking pauses occur between utterances so there are times when the circuit is idle. Measurements show that speech is present in each direction about 30 to 40 % of the time, averaged over a large number of busy trunks. A DCME exploits this low average channel activity to achieve a part of its reduction of needed channels.

Speech is encoded by ADPCM according to ITU-T Recommendation G.728 [81]; by 32 kbit/s for normal traffic loads, 24 or 16 kbit/s when the traffic load increases. (For 16 kbit/s the channel compression approaches 10:1, but the speech quality is then rather bad). During very heavy channel overloading, some voice channels may be subjected to a "freeze-out".

The DCME data detector discriminates between voiceband data and speech, so that higher-speed data signals are assigned to a ADPCM 40 kbit/s channel and protected from a "freeze-out". The DCME also discriminates between voiceband data and facsimile, using facsimile demodulation/remodulation processing.

The DCME can also pass a 64 kbit/s unrestricted (transparent) channel.

For more information about traffic handling, see table 1 of ITU-T Recommendation G.763 [82].

For traffic loads, consisting of various typical mixtures of speech, voiceband data, facsimile and unrestricted channels the DCME provides approximately a circuit multiplication of 5. (The actual circuit multiplication achieved will depend upon traffic loading, speech quality requirements, the percentage, as well as the type, of of voiceband data and facsimile, the number of on-demand unrestricted 64 kbit/s channels, the number of pre-assigned channels and the size of the interpolation pools).

The total delay associated with the dynamically assigned ADPCM encoded bearer channel by the transmit DCME shall not exceed 30 ms. The total delay associated with the dynamically assigned ADPCM bearer channel by the receive DCME shall not exceed 15 ms. (The delay values exclude the effects of Doppler and plesiochronous buffers and exclude the delays associated with the establishment and disestablishment of demand assigned 64 kbit/s unrestricted circuits. For facsimile demodulation/remodulation, the processing delay is given in ITU-T Recommendation G.766 [124]).

7.4 Mobile communication systems

7.4.1 General

ITU-T Recommendation G.174 [119] states that speech transmission performance achieved on mobile communication systems should be comparable to the PSTN/ISDN. Transmission performance criterias of wireless systems with influence on speech quality are listed below.

Connections including the wireless system should, under error-free conditions, achieve subjective ratings comparable to those of connections in the ordinary PSTN/ISDN. On the other hand to realize the widespread acceptance, the requirements to be satisfied by wireless systems may differ from the quality perspective. In addition, wireless systems should be robust to transcodings (mainly when used in tandem with a far-end wireless system), robust to a reasonable level of bit and frame errors, and robust to the wide variety of ambient noise conditions (e.g., offices, outdoors, motorways, cars) under which such systems will be used.

In order to reduce the number of speech channels on the radio link it is essential to adopt low bit-rate speech coding techniques. The speech coding algorithms cause delay and quantisation distortion. Propagation delay is also introduced by the radio channel coding algorithms. Radio channel coding is required to improve the error performance of the link. Reflections from stationary and non-stationary structural and environmental objects produce different propagation paths, with each path having a different time delay, phase and attenuation. Due to signal additions from the different (mulitpath) propagation paths, a portable radio channel will suffer signal fluctuations (i.e., mulitpath fading and dispersion) and distortions as a function of distance or time for a moving user. In fact, due to the motion of objects causing reflections in the vicinity of the user, even a stationary user will receive a signal that fluctuates with time. Interference from other users in adjacent cells may be experienced, particularly under fading conditions.

The quantization distortion of the different speech coding algorithms degrades the quality of speech. So far each algorithm has been designed to be unique to its mobile systems. Hence, when tandemed with other systems the quantization distortion has a cumulative effect.

As a result of the increased delay, echo control is required both to prevent echo from the mobile network to the PSTN/ISDN and from the PSTN/ISDN to the mobile network. This echo control is handled differently in the various mobile systems. Different systems have different standards for specifying the transmission performance from the mobile terminal to the point of interconnection with the fixed network. ITU-T Recommendations G.173 [37] and G.174 [119] offer general recommendations for public land mobile networks without referring to specific systems.

Mobile terminal equipment has to operate in unfavourable acoustic conditions. TBR 9 [134] specifies the European standard for mobile terminal equipment. However it does not adequately specify the ambient noise performance for mobile terminals. Therefore new specifications and testing methods are required in this field.

The design of the acoustic part of mobile terminals is critical to both ambient noise immunity and Terminal Coupling Loss (TCL). The latter determines whether echo control is neccessary in order to protect the PSTN/ISDN from echo. TBR 9 [134] also describes the sending and receiving sensitivity masks for mobile terminals. A practicable method for the user to counteract high ambient acoustic noise is to improve the SNR by increasing the receive gain (volume control). This requires a reduction of the send gain in order to prevent a reduction of the TCL, which would cause echo problems with the PSTN/ISDN.

The sidetone masking rating for mobile systems also originates in TBR 9 [134]. This rating affects the listener side tone rating (LSTR) which is critical in mobile systems.

7.4.2 Delay and echo

Echo control guidelines are necessary for the 'incremental delay' due to the wireless access portion of the connections to the PSTN/ISDN. Incremental delay is the delay added by the wireless system above and beyond that of the wireline segment being replaced.

An objective long-term value of less than 20 ms for one-way delay is desirable (see ITU-T Recommendation G.174 [119]). This delay refers to the duration of the one-way transmission between the wireless terminal and the base station including processing and propagation.

It is recommended that echo control measures are used when the one-way incremental delay per terminal end is 5 ms or higher.

The correlation between wireless systems which apply echo control measures and the wirebound network needs to consider some general aspects for planning purposes:

- echo control measures within the wirebound network which are not disabled in tandem with effects of the wireless system (e.g. transcoder, discontinous transmission (DTX), variations in echo response such as those due to handoffs that suddenly change the transmission path) can cause degradation of speech transmission quality;

- furthermore an uncontrolled correlation will arise with Digital Speech Interpolation (DSI) and DTX, because neither systems transmit speech pauses;

- also the DCME systems which are installed within the wirebound network can disturb the echo canceller within the wireless network.

Due to the amount of qdu's produced by wireless systems and the national and international limitation of inserted qdu's for a connection, PSTN/ISDN trunks with additional qdu's should be avoided. For connections between wireless systems and the wirebound network digital trunks without additional qdu's should be preferred.

No echo control is required for 4-wire all-digital connections, if the terminal coupling loss is higher than 46 dB. In interworking with PSTN/ISDN digital paths that terminate in analogue lines, the wireless system needs to provide echo cancellers.

In some configurations near-end echo may be coexistent with far-end echo. This may lead to an additional degradation in performance if the echo control device is unable to cope effectively with both echo signals.

7.4.3 Stability loss

For digital wireless access systems interfacing digitally to the PSTN/ISDN, a minimum loss of 6 dB between the digital input and output paths of the wireless system is recommended (see also subclause 6.5.6.1).

7.4.4 Quantization distortion

The long term objective is that the coder/decoder pair of the digital wireless system should not introduce a value of more than 4 qdu's, consistent with the long-term objective for Public Land Mobile Network (PLMN), as stated in ITU-T Recommendation G.173 [37]. It is recognised that in some cases, such as low bit-rate non-waveform coders, the measured qdu may not be appropriate, and other methods of specification should be used (see subclause 9.1.3.5 and annex G).

7.4.5 Clipping (temporal)

Speech clipping consists of four components: duration of clipping, percentage of speech clipped, frequency of clipping and overall speech activity.

7.4.6 Noise contrast and comfort noise

Noise contrast occurs when the background noise is interrupted. This may be caused by several effects, such as echo cancellation using center clippers, speech interpolation, discontinuous transmission, etc. Comfort noise will be introduced to mask the negative effects of noise contrast. Different systems may use different types of comfort noise. Recommendations on noise contrast limits as well as on comfort noise types and values are for future study.

7.4.7 Random bit errors and bursts of errors

Good speech quality should be maintained during up to 3 % frame erasures over any 10 s period (frame sizes of 10-20 ms are assumed). The usual criterion for speech quality is a reduction of no more than 0,5 Mean Opinion Score unit (5 point scale) relative to the error-free condition.

7.4.8 Speech coder performance

Speech coders for wireless personal communication including mobile communication need to be evaluated regarding their performance while being subjected to various lengths of correlated burst errors. Wireless communication systems are also likely to encounter higher ambient acoustic noise levels than wired communication systems. It is therefore also desirable that speech coders will be evaluated in different levels of acoustic background noise.

Of particular concern is the effect on the performance of radio channel fading while using low bit-rate voice coders. In this case wireless communication users may experience previously unknown types of degradation in voice quality resulting from the compression process itself or from error multiplication due to burst errors occurring in the compressed data. The impact on the perceived quality due to burst errors on low bit-rate speech coders is under study.

7.4.9 Examples of mobile communication systems

Analogue Cellular: TACS, NMT450, NMT900, RC2000, etc;

CT2: Cordless Telephone MKII, Common Air Interface (CAI) (I-ETS 300 131 [52]);

DAMPS: Digital Advanced Mobile Phone System (American digital standard IS 54 [69]);

DECT: Digital European Cordless Telephony (ETS 300 175 [47], TBR 10 [135]);

DCS 1800: Digital Communication System 1800 MHz (UK, Personal Communication Network (PCN));

DCT 900: Scandinavian Digital Cordless Standard;

FPLMTS (ITU-R): Future Public Land Mobile Telecommunications System;

GSM: Global System for Mobile Communication (I-ETS 300 020 [136], TBR 9 [134]);

TFTS: Terrestrial Flight Telephone System (ETS 300 326 [137]);

UMTS (ETSI): Universal Mobile Telecommunications System.

Table 10: Planning values of some mobile communication systems

System	One-way transmission time (ms)	Codec	Bit-rate (kbit/s)
CT2CAI	2,5	ADPCM	32
DAMPS	about 95	VSELP	8
DECT	14	ADPCM	32
DCS 1800	about 95	RPE-LTP	13
DCT 900	20	ADPCM	32
GSM	about 95	RPE-LTP	13
TFTS	about 95	MPLPC	9,6

7.5 ATM systems

7.5.1 General

Asynchronous Transfer Mode (ATM) is a packet switching technology that relays traffic via an address contained within the packet. Unlike other packet switching technologies such as ITU-T Recommendation X.25 [138] or Frame Relay (which are not used for voice transmission), ATM uses very short, fixed-length packets called cells. ATM represents a specific type of cell relay service defined as part of the overall B-ISDN standards. It can be based physically on the frame structure of SDH (Synchronous Digital Hierarchy).

The ATM cell is 53 bytes long, consisting of a 5 byte header containing the address and a fixed 48 byte information field. (Frame relay uses a 2 byte header and a variable length information field). The ATM cell is a compromise between the long frames generated by data communication applications and the short repetitive needs of voice.

ATM is a connection-oriented technique, i.e. a virtual connection must be established before the information is transmitted. This virtual connection remains during the whole connection time.

On a fibre cable with a length of 10 km there are 18 cells simultaneously on the way if the bit rate is 155 Mbit/s. Therefore it is not possible to introduce error correction during transmission. Any error correction has to be done by the terminals.

Up to now no special AAL-1 (ATM Adaptation Layer-1) specification for voice exists.

Table 11: Comparison between packet switching techniques

Switching technique	X.25	Frame Relay	ATM
Switching	Layer 3	Layer 2	Layer 1
Transmission rate	64 kbit/s	2 Mbit/s	155 Mbit/s
Error correction	very good	none	none
Protocol overhead	very large	low	large
Throughput	medium	large	very large
Protocol-transparency	none	good	yes
Bandwidth on demand	no	yes	yes
Dynamic routing	yes	no	no
Isochronous traffic	no	no	yes
Network delay	high	low	minor
International traffic	yes	no	no

7.5.2 Delay

Estimated values of the main overall delay components in ATM are shown in table 12:

Table 12: Estimated delays in ATM for voice transmission

Delay component	Delay (ms)
Packetization (of 64 kbit/s)	6,0
Cell delay variation compensation	2,0
ATM switching (4 * 0,25 ms)	1,0
SDH cross connecting (10 * 0,05 ms)	0,5
Propagation delay (1 000 km)	5,0
Exchange (2 * analogue/digital)	2,0
Total one-way delay	16,5

Comments:

Packetization delay should only occur once within a connection. A means of reducing this delay would be to send partly empty cells which however implies much less efficient information transfer. Thus half empty cells would decrease this type of delay to 3 ms. On the other hand, if low bit-rate traffic of 64/n kbit/s is routed via ATM with 100 % filled cells, the packetization delay increases to n*6 ms.

Cell delay variation compensation would become much larger if so-called shaping (to relax the effects of cell bursts) was applied on CBR (Constant Bit Rate) ATM connections. However, it seems to be generally agreed that shaping should not be used on CBR ATM connections.

The ATM switching delay per node as required by DBP Telecom and Bellcore is 0,25 ms. Four ATM switches can be considered as realistic for a national network.

SDH cross connecting adds a small amount of delay, about 0,05 ms per SDH cross connect. Ten SDH cross connects is a realistic figure for a national network. However, if the ATM backbone network (a meshed one) ends in local networks built up with ring structures the delay for the SDH cross connecting may increase up to 3 ms or more.

The propagation delay for fiber cables is 0,005 ms/km. The longest path within many countries (e.g. Germany) will be 1 000 km, leading to a delay of 5 ms.

The delay for both ends of the digital local exchange equipment is in total 2 ms.

7.5.3 Cell loss

The cell loss ratio is in the range of 10^{-9} to 10^{-6}. For 10^{-9}, this means that one cell in 70 days will be lost on the average if completely filled cells are used. For lost or inserted cells, 100 % filled, a break (or a dummy cell) of 6 ms in the PCM bit stream must be expected.

8 Combination effects of impairments on speech communication quality - background to the ETSI model

8.1 Introduction

The voice transmission quality is affected by a number of transmission parameters. Individual limits for acceptable performance are stated in ITU-T documentation for most of the important parameters.

However, combination effects of transmission parameters must also be considered in transmission planning. This is done in the computation models described in CCITT Supplement 3 [43], viz. the Transmission Rating (Bellcore), the CATNAP (UK), the Information Index (France) and the OPINE (NTT, Japan). Also, annex A to ITU-T Recommendation P.11 [26] presents a computation model, the Transmission Quality Index, TQI.

In ITU-T SG 12 these models have been used to evaluate some typical transmission examples and the computed results were fairly similar to each other.

The Transmission Rating model has been widely used as it is adopted as a standard in North America, albeit with parameters partly different from those used in CCITT.

There are several reasons why the ETSI computation model has been introduced. One is that the transmission planner today needs to take account of some new, important impairments that are not included in the previous models. Another is that recent subjective test results can be included more easily in the new model than by transforming and up-dating the old ones. Also, various "good" features "borrowed" from the different models could be incorporated in the ETSI model.

8.2 Derivation of the ETSI model

The ETSI model is derived from the information given in CCITT Supplement 3 [43] about models and from some recent CCITT reported results. (See subclauses 9.2.1 and 9.2.2 for a more detailed information).

The fundamental principle of the ETSI model is based on a concept given in the description of OPINE:

> **"Psychological factors on the psychological scale are additive."**

Thus, OPINE derives "Mean Opinion Scores", MOS, from a factor P of the form

$$P = Po - OPI \qquad (8.2.1)$$

where

$$OPI = PI(EL) + PI(N) + PI(AD) + PI(EC) + PI(ST) \qquad (8.2.2)$$

Po refers to the case with "no degradation".

The *PI*s are named "performance indexes" for various impairments, namely

PI(EL) = effective loudness loss or excess in speech;
PI(N) = noisiness during speech intervals and non-speech intervals;
PI(AD) = speech distortion for attenuation/frequency distortion;
PI(EC) = degradation caused by talker echo;
PI(ST) = degradation caused by sidetone.

The MOS values are obtained from the *P* factor by a mathematical procedure involving the Gaussian error (normal) distribution function.

The Bellcore Transmission Rating (BcTR) model is older than OPINE and is presented differently. However, a mathematical analysis shows that it can be given a very similar structure to the OPINE, i.e. the effects of impairments can be written as terms subtracted from a basic rating value. Also, estimates of users' opinions are derived by means of the normal distribution function.

Thus, it appears that the basic principle of additivity for psychological impairments is correct.

The ETSI model uses the same basic concept. For a connection, the transmission parameters are combined into a "Transmission Rating Factor" *R* which is transposed into estimates of users' satisfaction or dissatisfaction, i.e. percentages of opinions and/or Mean Opinion Scores (MOS). The user's degree of satisfaction or dissatisfaction relates both to his talking and listening over the connection. Note that the model will give separate estimations for how the users at either end of the connection perceive the transmission quality.

The Transmission Rating factor *R* is composed of the terms

$$R = Ro - Is - Id - Ie + A \qquad (8.2.3)$$

Ro represents in principle the basic SNR of the voice transmission at a 0 dBr point.

Is represents impairments which occur more or less simultaneously with the voice signal, such as those caused by too loud received speech level (OLR too small), non-optimum talker sidetone (STMR), quantizing distortion, etc.

Id represents impairments which appear delayed with regard to the voice signal, such as talker echo (TELR), listener echo (WEPL) and communication difficulties caused by too long absolute delay (Ta).

Ie represents transmission impairments caused by special equipment in the connection, such as certain low bit-rate codecs, Digital Circuit Multiplication Equipment (DCME), Voice Packeting Equipment (VPE), etc. (The effect of such equipment on speech quality is so complex that it is difficult to analyze by means of individual parameters*).*

A is termed *"Expectation Factor"*. It represents an "advantage-of-access" that certain systems have over conventional wirebound communication systems. The concept of "high quality" is intimately connected with how customers' expectations are fulfilled. (In this context, see also the general discussion of perceived voice transmission quality in clause 5*).* Thus, the overall connection transmission quality as perceived by the user is highly influenced by the ease or difficulty to establish a connection. In certain circumstances, wireless systems have an advantage that can compensate the subjective defects of some speech transmission effects. Examples are mobile telephony and multi-hop satellite connections to hard-to-reach regions. (Sometimes, for economical reasons, special *low-cost* long distance connections might be considered as having a similar advantage factor).

One may wish to make a direct comparison between the voice transmission quality of different systems. In this case the expectation factor *A* is simply omitted.

Is, Id and *Ie* are in turn formed as sums of separate impairment factors for OLR, STMR, qdu, TELR etc.

In many cases, only a single transmission parameter is the main determining factor of an impairment factor *I*. However, there are also impairment factors dependent on several parameters where certain masking effects are noticeable. (For instance echoes may be masked by circuit noise).

The expressions for *Ro* and the different *Is* in the ETSI model are largely derived from the BcTR model. The main reasons why are that the BcTR model is fairly simple and uses transmission parameters that (in principle) are available to network operators. Moreover, the BcTR model has been widely used when planning networks in North America. However, in some instances results from CATNAP or recent CCITT investigations have been used instead. Thus, the factors *Ie* and *A* are unique for the ETSI model.

The models in CCITT Supplement 3 [43] do not account for all impairments that can occur in a modern network.

However, the structure of the ETSI model makes it fairly simple to incorporate a new impairment, at least provisionally, as shown by the following examples.

The effect of an impairment is checked by subjective MOS tests under controlled conditions. The MOS values for "degradation" and "no degradation" are transformed into R-values, the difference of which correspond to the impairment factor *I*.

Alternatively, the degree of actual users' complaints (or the absence of such) caused by a particular impairment can be translated into an *R*-value which in turn gives the *I*-value. (This latter method can also be used to check that the model gives reasonable results).

Of course, the introduction of a new impairment factor has to be made with caution, preferably with the results being double-checked.

Note that the influence of any differences in the voice channel attenuation/frequency response between connections is not at present considered in the ETSI model. The model assumes terminals with good frequency responses and connections with essentially flat frequency responses. The reason is that with today's de-regulation and liberalization of terminal equipment the transmission planner may have difficulty in knowing the actual response curve for a particular user's channel, including the response of the telephone set, with sufficient accuracy to make a meaningful estimation of this influence on the speech communication quality. (It is assumed in the model that the frequency response is "within normal limits". Reference [103] gives an overview of how users perceive the influence of different frequency respones on various aspects of telephone speech quality, such as "total impression", "intelligibility", "naturalness", "brightness", "fullness", "clarity").

Other impairments, not yet included in the ETSI model, are for instance poor radio channel transmission for mobile communication and the influence of impulsive noise in normal wirebound networks. (Regarding the latter, some information can be had from ITU-T Recommendation P.55 [111] and note 2 of subclause 2.3.2 of ITU-T Recommendation P.11 [26]).

In what follows, the ETSI model will be discussed in more detail regarding the users' reactions as well as the rating and impairment factors. However, no elaborate equations will be given here as these can be found in subclause 9.1 "Description of algorithms in the ETSI model".

8.3 Estimates of users' reactions given in the ETSI model

Users' reactions to the voice transmission quality of a connection can be assessed in many ways, all of which are rather costly.

a) by user surveys, i.e. actual network users are interviewed and asked for their opinions. Results may be given as percentage of users finding the connection "good or better", GOB, or "poor or worse", POW;

b) by observing users' calling behaviour such as the percentages who (i) terminate their calls unusually early, (ii) re-dial or even (iii) actually complain to the network operator;

c) by subjective tests, involving test teams in controlled laboratory experiments. The results most often are given as "Mean Opinion Scores", MOS. (Sometimes also "percentage having difficulties" are recorded).

The aim of computation models is to give estimates of how such user assessments would turn out, provided the assessments really were made. One must note, however, that users' opinions vary with time and circumstances so that there is an inherent uncertainty anyway. Thus, there is in principle a limit to the "precision" of any model. The advantage of using a model is that transmission planning and comparisons of transmission conditions can be done in a systematic and consistent way.

In the ETSI model, the following methodology has been applied for the estimates of user reactions:

For estimates of category a), GOB and POW are obtained from the *R*-factor by means of the normal distribution function

$$E(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt \qquad (8.3.1)$$

Thus

$$GOB = 100E[(R - 60) / 16]\ \% \qquad (8.3.2)$$

$$POW = 100E[(45 - R) / 16]\ \% \qquad (8.3.3)$$

The expressions for GOB and POW are chosen as approximate mean values of the curves given in CCITT Supplement 3 [43], figure 1-6, for the Bellcore Transmission Rating model, namely "Long toll", "CCITT" and "MH" (see figure 53).

For network operators, estimates of category b) would normally be of most interest but the factual basis for these are also rather difficult to obtain. In the ETSI model, the percentage of users "terminating calls early", TME, is given as

$$TME = 100E[(36 - R) / 16]\ \% \qquad (8.3.4)$$

This relation is based on results from "AT&T Long toll" interviews as cited in J.Gruber, G.Williams: "Transmission Performance of Evolving Telecommunications Networks", p. 34 [53].

Finally, in the ETSI model the expression for MOS as function of R has been derived indirectly, partly by comparison with relations between GOB, POW and MOS as given in Annex A to ITU-T Recommendation P.11 [26], partly from experience of what the range of scores is that subjective test teams usually assign to various speech transmission channels. A direct conversion from ITU-T Recommendation P.11 [26] gives the range 1 < MOS < 5. However, in actual subjective tests the maximum MOS is hardly ever higher than 4,5 and therefore the ETSI model has been adapted to this value by a linear transformation from the ITU-T Recommendation P.11 [26] results. Thus, for the ETSI model the following algorithm applies

For $0 < R < 100$ $\quad MOS = 1 + 0,035\,R + R\,(R - 60)\,(100 - R)\,7\,10^{-6}$ (8.3.5)

For $R < 0$ $\quad MOS = 1$ (8.3.6)

For $R > 100$ $\quad MOS = 4,5$ (8.3.7)

The relation between *R* and MOS is an almost linear, slightly S-shaped curve (see figure 40).

8.4 Discussion of ratings and impairment factors in the ETSI model

NOTE: For detailed equations, see subclause 9.1 "Description of algorithms in the ETSI model".

The basic form of the ETSI *R*-factor is given in equ.(8.2.3). The expression for *Ro*, the "basic signal-to-noise ratio", has been derived from the Bellcore Transmission Rating (BcTR) model in the following way:

The BcTR model gives *R* as a function of OLR and the total noise *Nf* dBmp "at a point where RLR = 0 dB". A numerical analysis shows that this *R*-factor can, with quite good accuracy, be written as the difference between two terms. The first is a linear function of (OLR + *Nf*), while the second only appears in the form of a correction term when OLR is low. i.e. the first term can be interpreted as a "basic signal-to-noise ratio" and the second as an impairment factor *Iolr* caused by a too loud connection.

In the ETSI model, the total noise *No* is referred to a 0 dBr point for practical reasons and thus the "basic signal-to-noise ratio" *Ro* is a function of SLR instead of OLR.

$$Ro = 15 = 1,5(SLR + No) \qquad (8.4.1)$$

No dBm0p is the sum of all electric circuit noises and the equivalent circuit noises caused by room noise pick-up at both ends of the connection. (The room noise algorithms are derived from the BcTR model and some recent investigations, published in CCITT and ETSI).

There is also a "noise floor" *Nfor* incorporated in *No*, corresponding to the fact that the human ear is not affected by noises below a certain level.

Note that quantizing distortion noise is not included in *No*.

The "simultaneous-impairment" factor *Is* is

$$Is = Iolr + Ist + Iq \tag{8.4.2}$$

Iolr represents the decrease in quality caused by a too loud connection, i.e. OLR is too low.

Ist represents the impairment caused by non-optimum sidetone.

Iq represents the impairment caused by quantizing distortion from PCM and ADPCM codecs, digital pads, etc.

The ETSI impairment factor *Iolr* is a function of OLR but also, in contrast to what the BcTR model implies, a function of the total noise perceived by the listener. The explanation is that at high noise levels a low OLR is an advantage rather than a disadvantage, a fact verified by subjective tests and results from other models.

The expression for *Ist* is derived from the general experience of talker sidetone as discussed in Supplement 11 to the P-series Recommendations, [41]. (i.e. there is a non-critical mid-range of STMR giving no impairment, while a very low STMR value is worse than a very high, see figure 45). The mathematical function for *Ist* is not very critical.

The expression for *Iq* is derived from CATNAP in which model the impairment caused by quantizing distortion is compared with the effect of an equivalent noise. (The BcTR model uses a similar approach but the result of CATNAP appears to be more realistic).

The "delayed-impairment" factor *Id* is

$$Id = Idte + Idle + Idd \tag{8.4.3}$$

Idte and *Idle,* the impairment factors for talker and listener echo respectively, are derived from the BcTR model in the form of changes in *R* as function of echo losses and delay times. There is also a masking effect of noise. For *Idte,* additional information from CCITT Contributions has been used.

Idd is estimated from the information given in ITU-T Recommendation G.114 [36], about the difficulties encountered with long absolute delays even with perfect echo cancellation. (The limit *Ta* = 400 ms in ITU-T Recommendation G.114 [36] corresponds to *Idd =* 21, a tolerable impairment. *Ta* = 800 ms gives *Idd =* 42, a quite serious impairment).

The "equipment-impairment" factor *Ie* has been introduced as a flexible mean to encompass impairments caused by more complicated, new equipment. For natural reasons, the uncertainty of *Ie*-values may be rather large until more field experience is gained of the equipment in question.

The "expectation" factor *A* has been introduced in response to users' favourable reactions to mobile services even when those have less than optimum speech quality compared with wirebound telephony.

Finally, some comments will be given on those parameters and impairments which are *not* included in the ETSI model although they are in other models.

Electro-acoustic response characteristics of handsets are not incorporated in the ETSI model because with the present liberalization and proliferation of type-approved sets the network operators cannot possibly know which handset a customer uses.

The ETSI model does not consider attenuation/frequency-distortion in the electric path. The reason why is that such distortion is much less in magnitude than the electro-acoustic loss variations due to the user's handling of his set as well as the difference in response curves between types of telephone sets.

8.5 Future up-dating and amendments of the ETSI model

It seems very desirable to include effects of impulsive and time-varying noise. Also, time-varying delays as may occur in modern ATM systems must be considered eventually. Annex M lists some specific points which require further studies.

9 A computation model for the overall voice transmission quality from mouth to ear

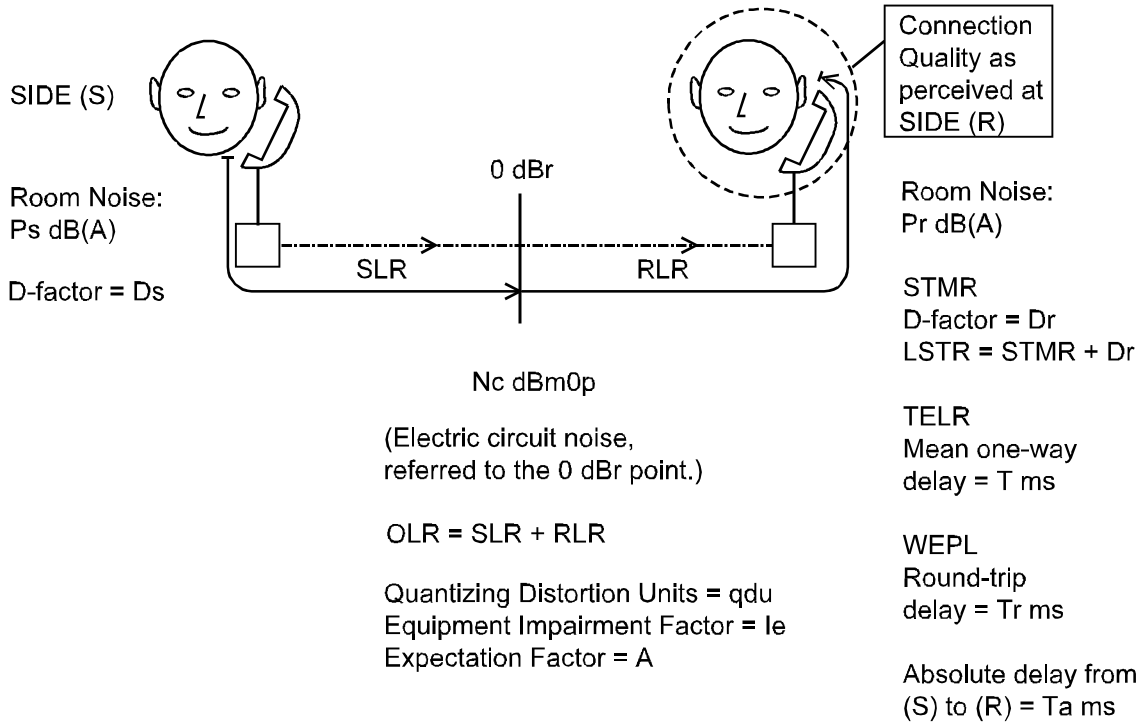
9.1 Description of algorithms in the ETSI model

9.1.1 Introduction

The model provides "voice transmission quality" measures for handset telephony in the normal speech band of nominally 300 Hz to 3 400 Hz.

The configuration of the connection and its relevant transmission parameters is depicted in figure 38.

The model estimates the speech communication quality mouth to ear as perceived by the user at side R, both as listener and talker. If the connection is un-symmetrical, the speech communication should of course also be evaluated in the other direction.



NOTE: Evaluation of the voice transmission quality (R-factor, GOB, POW etc) refers to the user at side R, both as listener and talker.

Figure 38: Configuration of the connection and the relevant transmission parameters

9.1.2 "Quality measures" provided by the ETSI model

The transmission parameters are combined into a "*Transmission Rating Factor" R* which can lie in the range from 0 to 100 or even higher. $R = 0$ represents an extremely bad and $R = 100$ a very high transmission quality of the connection.

The *R*-factor is transposed into a number of different "quality measures" representing statistical estimations of:

a) percentage of users finding the connection
"Good or better" = GOB;
"Poor or worse" = POW;

b) percentage of users terminating their calls early due to bad transmission = TME;

c) Mean Opinion Score = MOS (Scale 1 - 5).

Estimations according to category a) approximately simulates results obtained in user surveys, i.e. when users are interviewed and directly asked for their opinion.

Category b) indicates users' spontaneous (mildest) reaction to bad transmission quality. (Note that in general the percentage TME < POW).

Category c) simulates results obtained by subjective tests made under controlled conditions.

GOB, POW and TME are obtained from the R-factor by means of the Gaussian error function

$$E(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt \qquad (9.1.1)$$

Thus

$$GOB = 100 E\left(\frac{R-60}{16}\right) \quad \% \qquad (9.1.2)$$

$$POW = 100 E\left(\frac{45-R}{16}\right) \quad \% \qquad (9.1.3)$$

$$TME = 100 E\left(\frac{36-R}{16}\right) \quad \% \qquad (9.1.4)$$

GOB, POW and TME as a function of the R-factor are depicted in figure 39.

MOS, scale 1 - 5, is obtained directly from the R-factor by the expressions:

For $0 < R < 100$ $\quad MOS = 1 + 0,035\,R + R\,(R - 60)\,(100 - R)\,7\;10^{-6}$ (9.1.5)

For $R < 0$ $\quad MOS = 1$ (9.1.6)

For $R > 100$ $\quad MOS = 4,5$ (9.1.7)

MOS as function of R is depicted in figure 40.

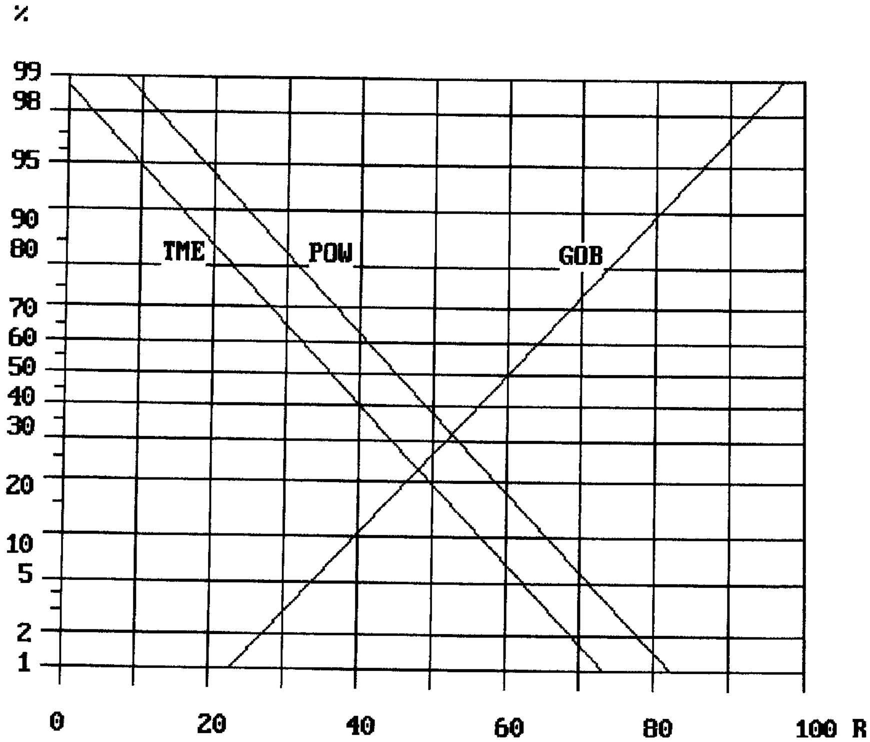
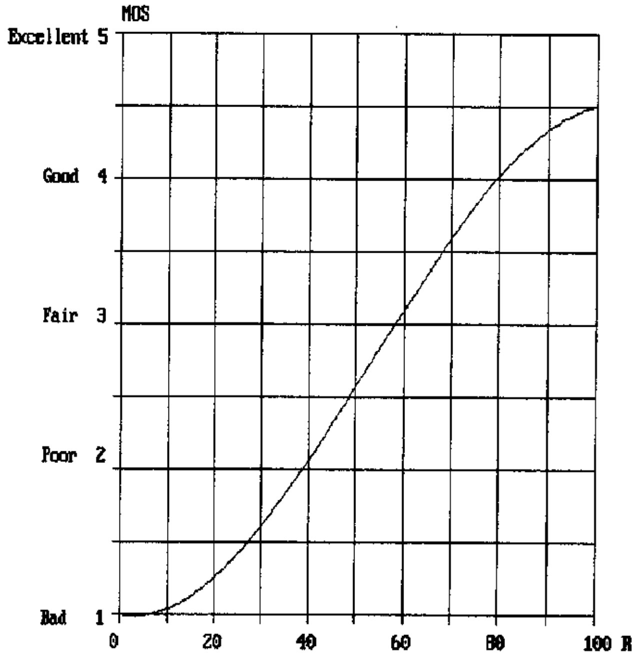


Figure 39: Percentages GOB (Good Or Better), POW (Poor Or Worse) and TME (Terminate calls early) as function of rating factor R



NOTE: MOS values, both predicted and subjectively measured, need always to be interpreted with care. In general, subjective MOS values vary between test occasions.

Figure 40: Calculated (predicted) MOS as function of rating factor R

9.1.3 The transmission rating factor R

9.1.3.1 General

The transmission rating factor *R* is composed of the terms

$$R - Ro - Is - Id - Ie + A \qquad (9.1.8)$$

Ro represents in principle the basic SNR of the voice transmission at the 0 dBr point nearest Side (R).

NOTE 1: A connection may consist of several cascaded "circuits" (in the CCITT sense), each circuit having its own 0 dBr transmission reference point. For certain network configurations it may be necessary to introduce so-called "level jumps" at the interface between two circuits, i.e. there may exist a loss or gain between two different 0 dBr transmission reference points!

Is represents impairments which occur more or less simultaneously with the voice signal, such as caused by too loud received speech level (OLR too small), non-optimum sidetone (STMR), quantizing noise (qdu), etc.

Id represents impairments which appear delayed with regard to the voice signal, such as talker echo (TELR), listener echo (WEPL) and communication difficulties caused by too long absolute delay (Ta), etc.

Ie represents transmission impairments caused by special equipment in the connection, such as certain low bit-rate codecs, DCME, VPE, etc. (The effect of such equipment on speech quality is so complex that it is difficult to analyze by means of individual parameters).

NOTE 2: The individual transmission parameters can appear in more than one of the *I*-factors, sometimes in a masking effect.

A is termed *"Expectation Factor"*. It represents an "advantage-of-access" that certain systems have over conventional wirebound communication systems. The concept of "high quality" is intimately connected with how customers' expectations are fulfilled. Thus, the overall connection transmission quality as perceived by the user is highly influenced by the ease or difficulty to establish a connection. In certain circumstances, wireless systems have an advantage in this respect over wire systems, an advantage that can compensate the subjective effect of some speech transmission defects. Examples are mobile telephony and multi-hop satellite connections to hard-to-reach regions. (Sometimes, for economical reasons, special *low-cost* long distance connections might be considered as having a similar advantage factor).

9.1.3.2 Noise considerations and the basic "signal-to-noise" factor *Ro*

The expression for *Ro* is

$$Ro = 15 = 1{,}5(SLR + No) \qquad (9.1.9)$$

SLR is referred to the 0 dBr point nearest Side (R).

No is the *total* noise in dBm0p, also referred to the 0 dBr point. *No* is obtained by power addition of

- the electric circuit noise *Nc* dBm0p;
- the equivalent *circuit* noise *Nos* dBm0p caused by the *room* noise *Pos* dB(A) at Side (S);
- the equivalent *circuit* noise *Nor* dBm0p caused by the *room* noise *Por* dB(A) at Side (R);
- the "Noise Floor" *Nfo* dBm0p caused by conditions at side (R).

The electric circuit noise *Nc* is obtained by power addition of the various electric noise sources in the connection, all referred to the 0 dBr point. (If a noise source of *N* dBmp is introduced at a point of relative level *L* dBr in the circuit, this corresponds to a noise level of (*N + L)* dBm0p at the 0 dBr point).

The equivalent circuit noise *Nos* caused by room noise *Pos* dB(A) at Side (S) is

$$Nos = Pos = SLR - Ds - 100 + 0{,}008(Pos - OLR - Ds - 14)^2 \quad \text{dBm0p} \qquad (9.1.10)$$

where

$$OLR = SLR + RLR \qquad (9.1.11)$$

Ds is the D-factor of the handset at side (S). The influence of room noise *Pos* at side (S) is depicted in figure 41.

The equivalent circuit noise *Nor* caused by room noise *Por* dB(A) at Side (R) is

$$Nor = RLR - 121 + Pore + 0{,}008(Pore - 35)^2 \quad \text{dBm0p} \qquad (9.1.12)$$

where *Pore* is the *effective* room noise caused by enhancement of *Por* by the listener's sidetone path

$$Pore = Por + 101g[1 + 10^{(10\text{-LSTR})/10}] \text{ dBm0p} \qquad (9.1.13)$$

LSTR is the Listener's Sidetone Rating at side (R). The enhancement of room noise *Por* to *Pore* due to LSTR at side (R) is depicted in figure 42. The equivalent circuit noise *Nor* caused by the effective room noise at side (R) is depicted in figure 43.

The noise floor *Nfo* dBm0p refers to a noise floor *Nfor* dBmp at Side (R). Thus

$$Nfo = Nfo = Nfor + RLR \qquad \text{dBm0p (9.1.14)}$$

Normally, *Nfor* = -64 dBmp.

Finally**, the total noise *No*** is obtained by power addition of the noise components:

$$No = 101\text{g}[10^{Nc/10} + 10^{Nos/10} + 10^{Nor/10} + 10^{Nfo/10}] \text{ dBm0p} \qquad (9.1.15)$$

NOTE: Quantizing distortion noise is *not* included in this summation.

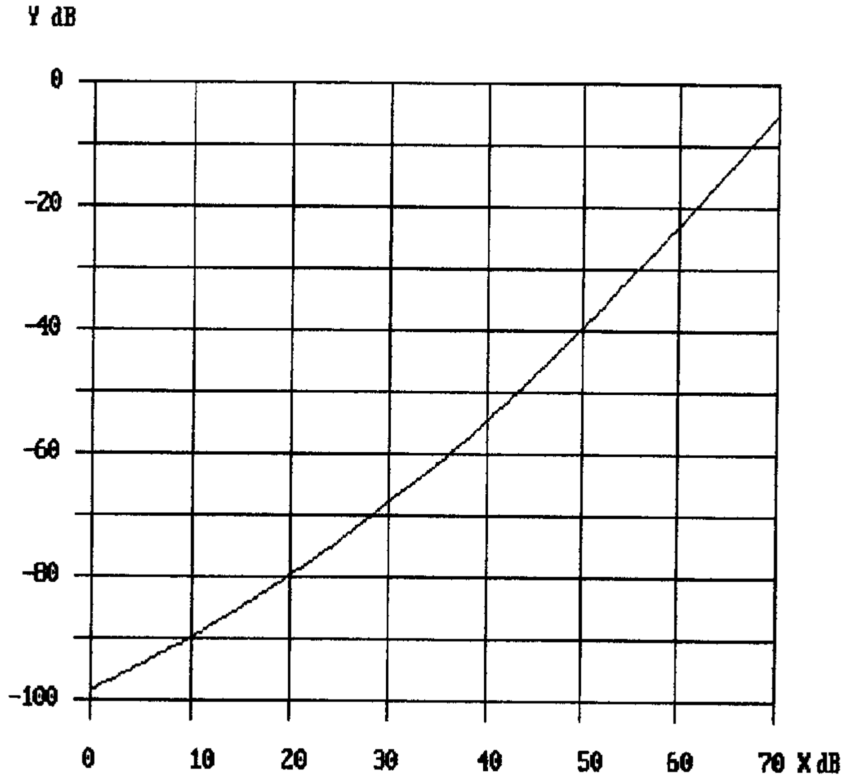


Figure 41: Influence of room noise Pos dB(A) at side S. Equivalent circuit noise Nos = RLR + Y dBm0p; X = Pos - OLR - Ds

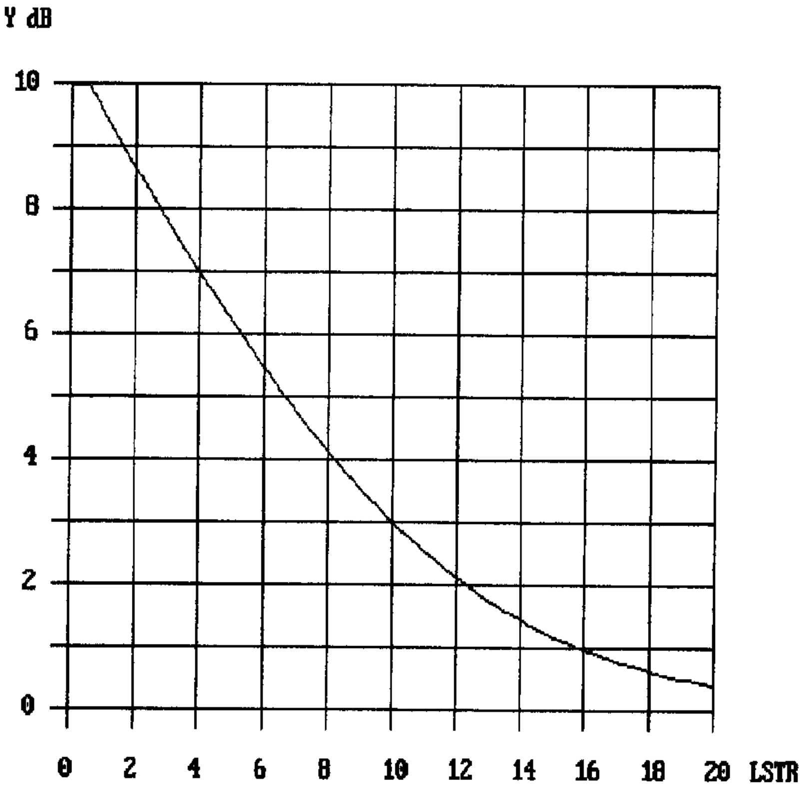


Figure 42: Enhancement of room noise Por to Pore due to LSTR at side R. Pore = Por + Y dB(A)

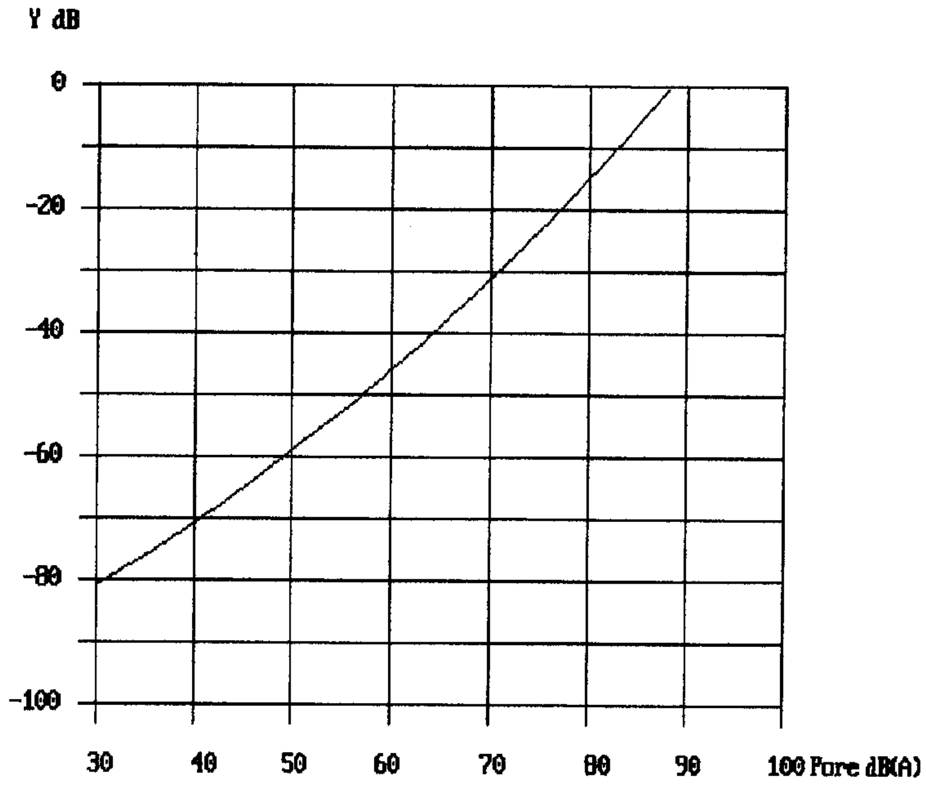


Figure 43: Equivalent circuit noise Nor dBm0p caused by the effective room noise at side R. Nor = RLR + Y dBm0p

9.1.3.3 The "simultaneous-impairment" factor *Is*

The expression for *Is* is

$$Is = Iolr + Ist + Iq \tag{9.1.16}$$

Iolr represents the decrease in quality caused by a too loud connection, i.e. when *OLR* is too low.

Ist represents the impairment caused by non-optimum sidetone.

Iq represents the impairment caused by quantizing distortion from PCM, digital pads, etc. Distortion caused by ADPCM and low bit-rate codecs is covered by the impairment factor *Ie* (see subclause 9.1.3.5).

The expression for *Iolr* is

$$Iolr = 20\left[\left\{1 + (X/8)^8\right\}^{1/8} - X/8\right] \tag{9.1.17}$$

where

$$X = OLR + 0{,}2(64 + Nt) \tag{9.1.18}$$

$$Nt = No - RLR$$

The impairment factor *Iolr* as function of OLR is depicted in figure 44.

The expression for *Ist* is

$$Ist = 10 \cdot \left[1 + \left\{(STMRo - 12)/5\right\}^6\right]^{1/6} - 46 \cdot \left[1 + \left\{STMRo/23\right\}^{10}\right]^{1/10} + 36 \tag{9.1.19}$$

where

$$STMRo = -10 \cdot \lg\left[10^{-STMR/10} + e^{-T/4} \cdot 10^{-TELR/10}\right] \qquad (9.1.20)$$

The impairment factor *Ist* as function of STMR is depicted in figure 45.

The expression for *Iq* is

$$Iq = 15 \lg\left[1 + 10^{Y}\right] \qquad (9.1.21)$$

where

$$Y = (Ro - 100)/15 + (46 - G)/10 \qquad (9.1.22)$$

Ro is given by equ.(9.1.9) and *G* by the expressions

$$G = 1,07 + 0,258Q + 0,0602Q^{2} \qquad (9.1.23)$$

$$Q = 37 - 15 \cdot \lg(qdu) \qquad (9.1.24)$$

qdu is the number of *quantizing distortion units* in the connection.

The impairment factor *Iq* as function of qdu is depicted in figure 46.

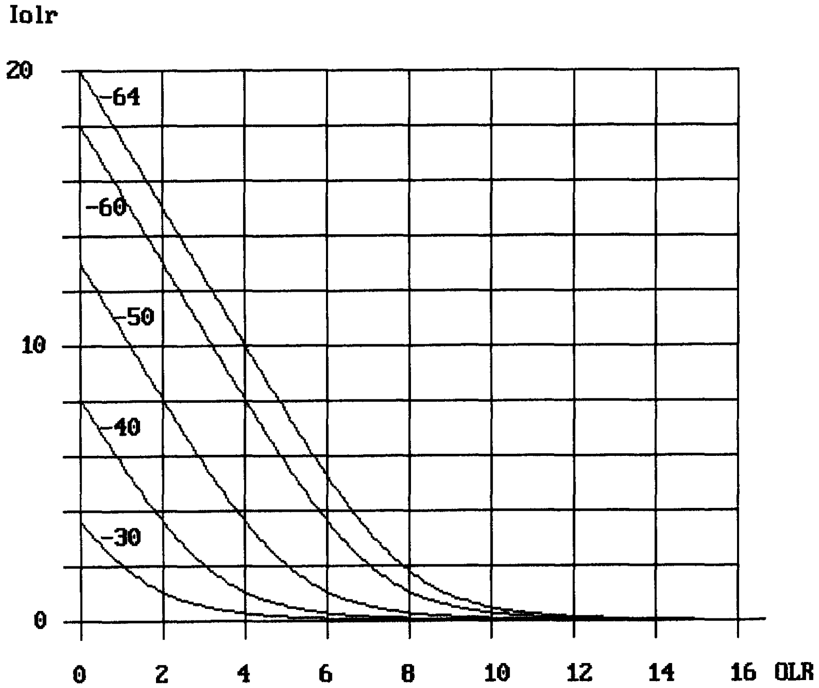


Figure 44: Impairment factor Iolr as function of OLR. Parameter for the curves is Nt = No - RLR dBmp; No = circuit noise (in dBm0p) at the 0 dBr point

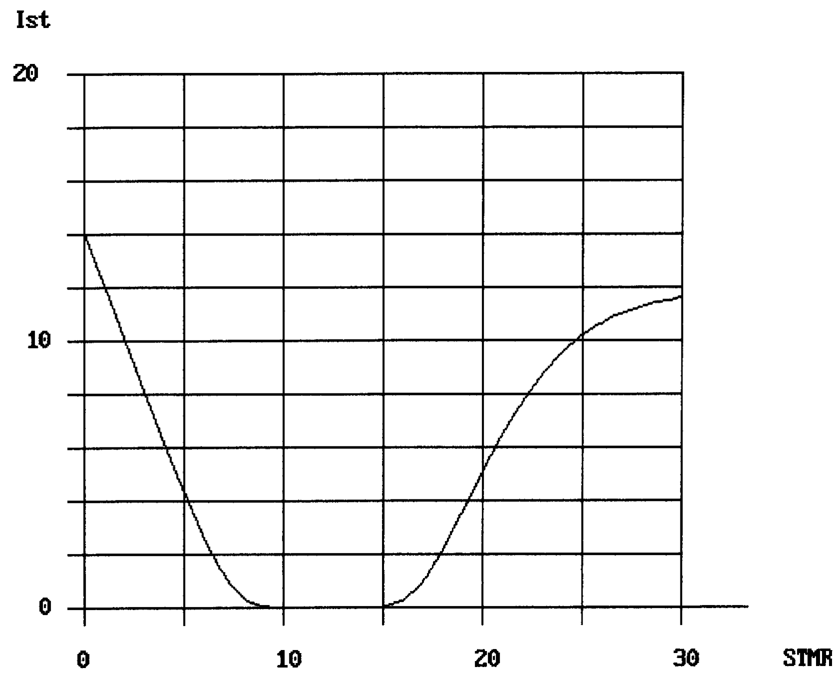


Figure 45: Impairment factor Ist as function of STMR

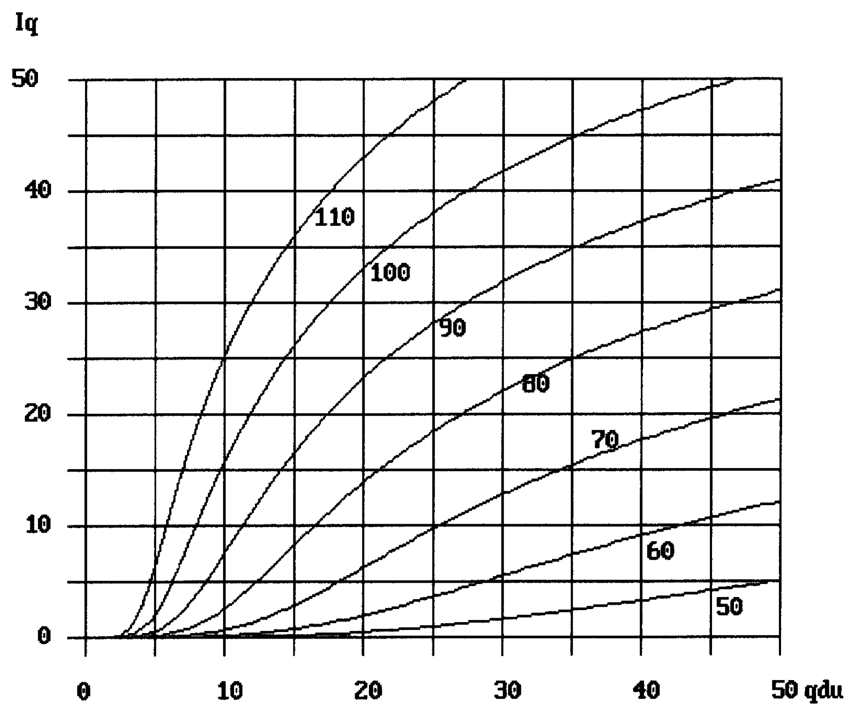


Figure 46: Impairment factor Iq as function of qdu. Curve parameter = (SLR + No) dBmp

9.1.3.4 The "delayed impairment" factor *Id*

The expression for *Id* is

$$Id = Idte + Idle + Idd \qquad (9.1.25)$$

Idte represents the impairment caused by talker echo. Relevant parameters are the Talker Echo Loudness Rating (*TELR)* and the mean one-way delay time *T* ms for the echo.

Idle represents the impairment caused by listener echo. Relevant parameters are the Weighted Echo Path Loss (WEPL) and the round-trip delay *Tr* ms for the echo.

Idd represents the impairment caused by too long absolute delay which occurs even with perfect echo cancelling. Relevant parameter is the one-way absolute delay *Ta* ms.

The expression for *Idte* is

$$Idte = \left[(Roe - \mathrm{Re}) / 2 + \sqrt{(Roe - \mathrm{Re})^2 / 4 + 100} - 1 \right] \cdot (1 - e^{-T}) \qquad (9.1.26)$$

where

$$Roe = -1{,}5 \cdot (No - RLR) \qquad (9.1.27)$$

$$\mathrm{Re} = 80 + 2{,}5(TERV - 14) \qquad (9.1.28)$$

$$TERV = TELR - 40 \lg \frac{1 + T/10}{1 + T/150} + 6e^{-0{,}3T^2} \qquad (9.1.29)$$

NOTE: For $T < 1$ ms the "talker echo" should be considered as *sidetone* so that then *Idte* = 0.

See figures 47 and 48 for relation between TELR and TERV.

Equ.(9.1.26) to (9.1.29) apply when the talker sidetone is "normal", i.e. 9 < STMR < 15. For lower values of STMR the talker echo is partly masked by the sidetone, while for higher values of STMR the talker echo becomes more noticeable than with a normal sidetone. These phenomena are taken into consideration by adjustments of the TERV and Idte respectively as follows:

For STMR < 9: TERV is replaced by TERVs in equ.(9.1.28)

$$TERVs = TERV + Ist / 2 \qquad (9.1.30)$$

For STMR > 15: Idte is replaced by Idtes

$$Idtes = \sqrt{Idte^2 + Ist^2} \qquad (9.1.31)$$

The impairment factor *Idte* as function of TERV is depicted in figure 49.

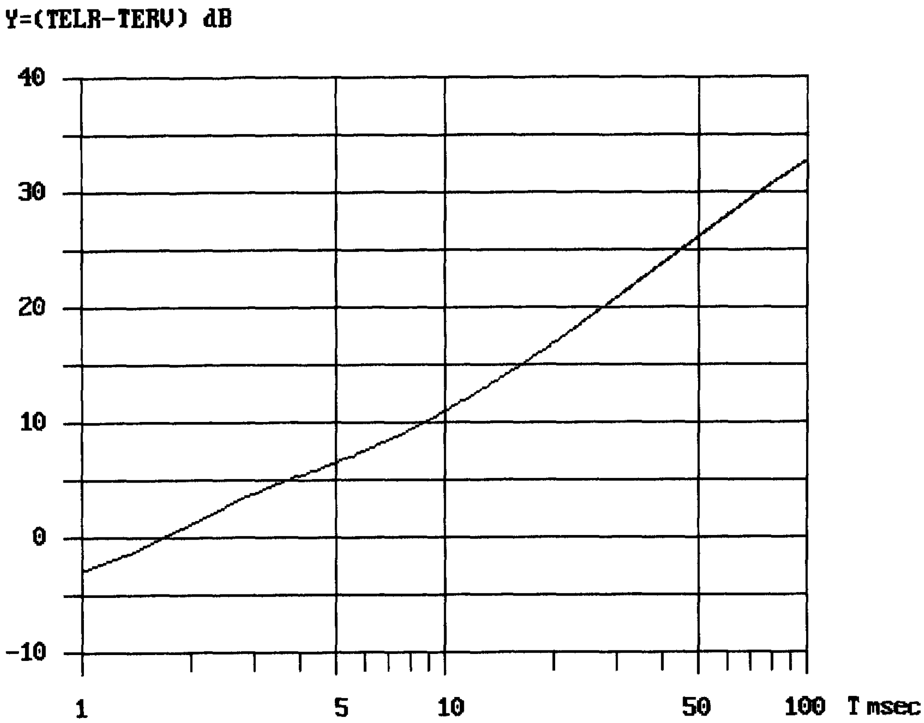


Figure 47: TERV = TELR - Y; T = mean one-way transmission time

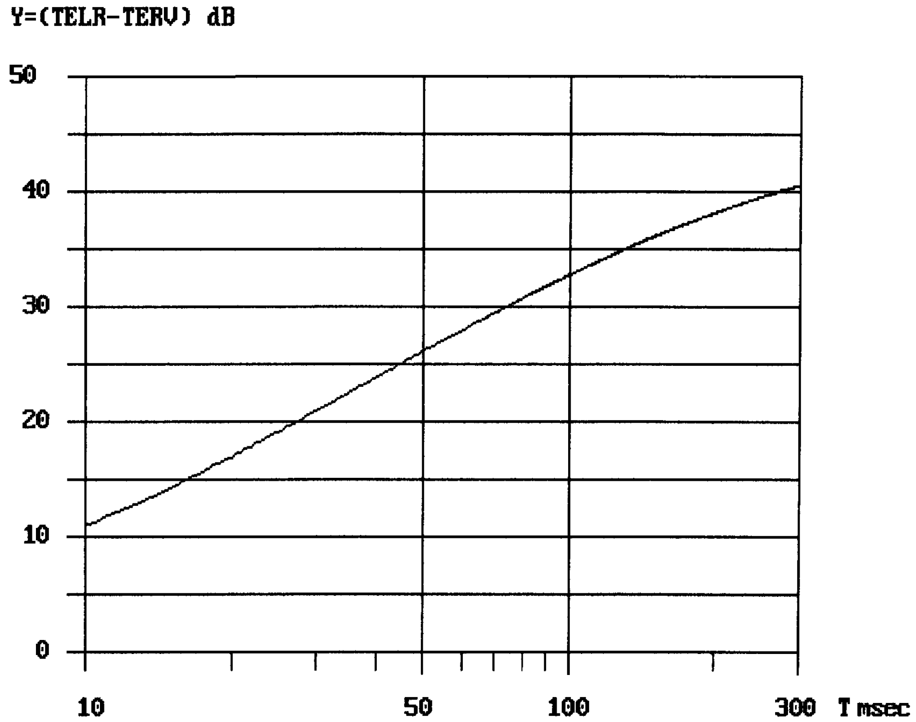
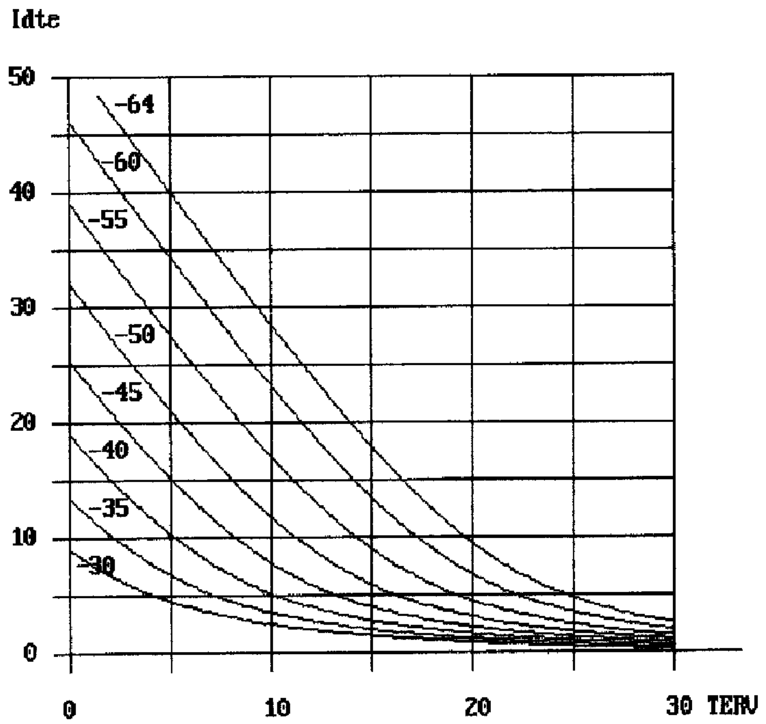


Figure 48: TERV = TELR - Y; T = mean one-way transmission time



NOTE: The curves are valid for T > 2 ms.

Figure 49: Impairment factor Idte as function of TERV; curve parameter = No - RLR dBmp

The expression for *Idle* is

$$Idle = (Ro - Rle)/2 + \sqrt{(Ro - Rle)^2/4 + 169} \quad (9.1.32)$$

where

Ro is given by equ.(9.1.9), and

$$Rle = 10{,}5(WEPL + 7)(Tr + 1)^{-0{,}25} \quad (9.1.33)$$

The parameter *Rle* as function of round-trip delay *Tr* is depicted in figure 50. The impairment factor *Idle* as function of parameter *Rle* is depicted in figure 51.

The expressions for *Idd* are

For $Ta < 100$ ms $\quad Idd = 0 \quad (9.1.34)$

For $Ta > 100$ ms:

$$Idd = 25.\left\{\left(1 + X^6\right)^{1/6} - 3.\left(1 + \left[X/3\right]^6\right)^{1/6} + 2\right\} \quad (9.1.35)$$

where

$$X = \frac{\lg(Ta/100)}{\lg 2} \quad (9.1.36)$$

The impairment factor *Idd* as function of the absolute one-way transmission time *Ta* is depicted in figure 52.

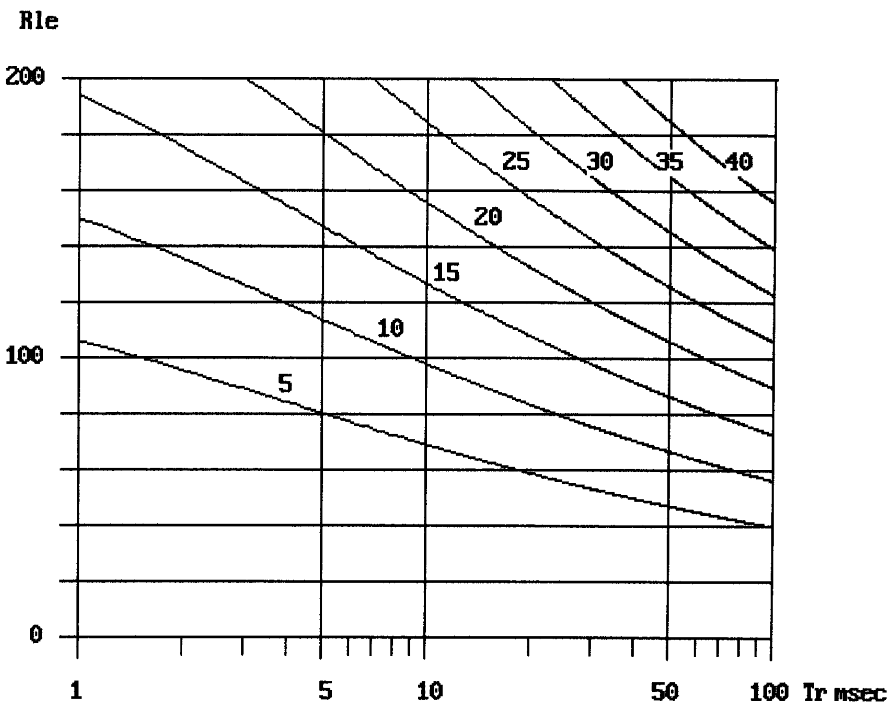


Figure 50: Parameter Rle as function of round-trip delay Tr; curve parameter = WEPL

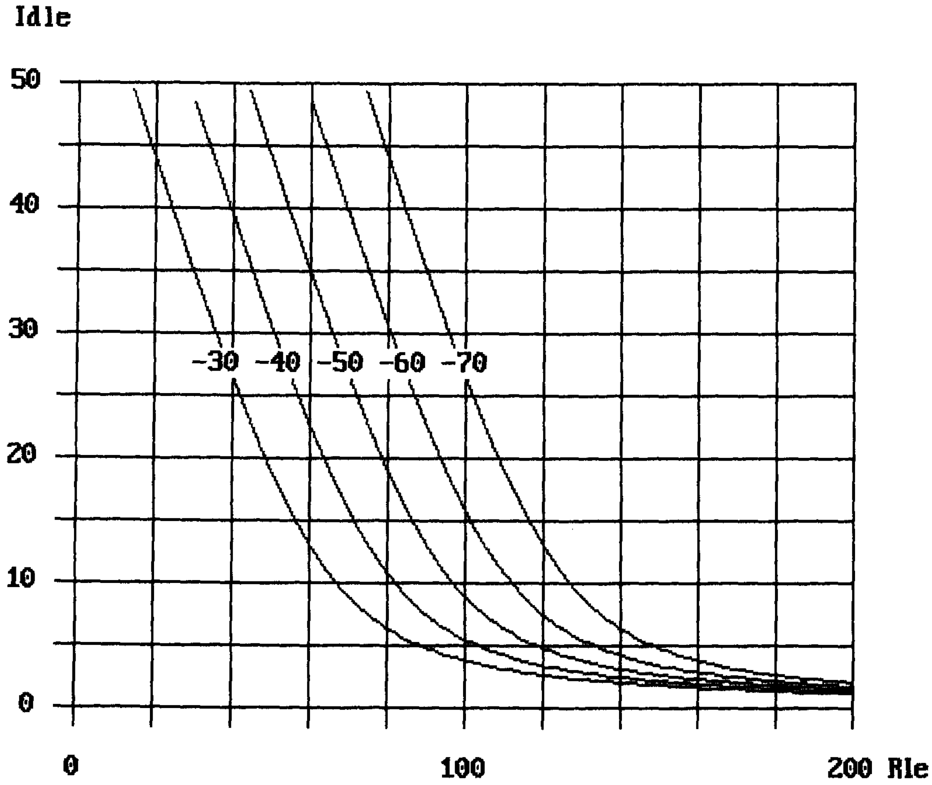


Figure 51: Impairment factor Idle as function of parameter Rle; curve parameter = (SLR + No) dBmp

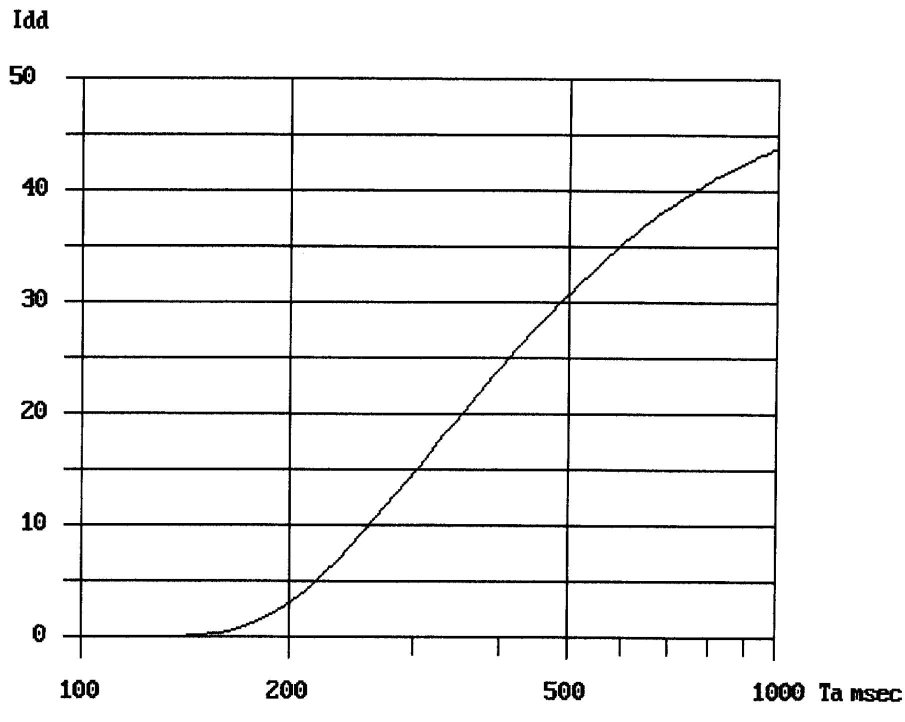


Figure 52: Impairment factor Idd as function of the absolute one-way transmission time Ta

9.1.3.5 The "equipment impairment" factor *Ie*

The value of *Ie* depends of course on test results for the particular equipment concerned as well as on network experience. Very much remains to be done. However, some examples with provisional results will be given in what follows.

Low bit-rate codecs in tandem:

A low bit-rate codec of a particular type is associated with a specific *equipment impairment factor Ie* which is additive to other impairment factors. Provisionally, the values given in table 13 can be applied for transmission planning.

Table 13: Equipment impairment factors *Ie* = *K* for low bit-rate codecs in transmission planning (Provisional)

Codec	kbit/s	*K*
ADPCM	40	2
	32	7
	24	25
	16	50
LD-CELP	16	7
	12,8	20
VSELP	8	20
RPE-LTP	13	20
CELP+	6,8	25

9.1.3.6 The "expectation" factor *A*

For the factor A, defined in subclause 8.2, the following provisional values are proposed:

- *A*= 0 for conventional (wirebound) systems;
- *A*= 5 for DECT and similar mobile systems;
- *A*= 10 for GSM and similar mobile systems;
- *A*= 20 for remote locations which only can be reached via multi-hop satellite connections.

9.2 Comparisons of results from the ETSI model, other models and subjective tests

9.2.1 Introduction

The fundamental principle of the ETSI model is based on a concept given in the description of the computation model OPINE:

> "Psychological factors on the psychological scale are additive".

This principle in turn appears to be based on some fundamental work of J. Allnatt some 20 years ago, see for instance reference [5].

The ETSI R-rating is in principle the same as the R-rating in the "North American Transmission Rating model". This is described in CCITT Supplement 3 [43] by Bellcore, however with ITU-T parameters for Loudness Ratings, etc. Here it will be called the BcTR model.

Although it is not explicitly stated in the description of the BcTR model it turns out by a mathematical analysis that the R-rating is such a "psychological scale" on which all impairments, dealt with in the BcTR model, can be transformed into additive "psychological factors". Thus, in view of the general additivity principle it is assumed, with a certain plausibility, that both the "old" and the "new" impairments included in the ETSI model obey the additivity law on the R - scale. (The mathematical proof of the additivity properties of the impairments in the BcTR model is quite simple and is left to the interested reader).

The BcTR model has been the main source for the ETSI model. The models presented in Annex A to ITU-T Recommendation P.11 [26] and CCITT Supplement 3 [43] have also been considered as well as the results from some recent subjective tests of talker echo, reported in the literature.

In particular, the following sources for the impairment evaluations can be mentioned:

- the influence of room noise at send and receive sides, including the dependence on LSTR and D - factors: the BcTR model and published test results from the Australian and Swedish Administrations;

- the influence of loudness ratings - OLR, SLR, RLR - and equivalent circuit noise: the BcTR model, however adjusted for low OLR values by considerations of results from the CATNAP model and some additional subjective tests;

- the influence of the talker sidetone, STMR: interpretation of information in CCITT Supplement 11 to the P-series Recommendations [41] and the information, that talker echo with very low delay is interpreted as a form of sidetone;

- the influence of quantizing distortions for PCM systems (qdu): the CATNAP (see CCITT Supplement 3 [43]) model;

- the influence of talker echo, TELR: For normal values of STMR the BcTR model has been used, slightly adjusted by considerations of results from NTT investigations, and complemented for short delays by results published by France Telecom and Telia Research. For very low values of STMR results from the BcTR model have been applied, while for very high values of STMR subjective tests from Telia Research and British Telecom have been employed;

- the influence of listener echo, WEPL: the BcTR model;

- the influence of long absolute delays: interpretation of information given in ITU-T Recommendation G.114 [36], with a certain emphasis on how the delay influences highly interactive conversations up to about 800 ms. For still longer delays, the conversation partners are likely to realize that one has to wait for an answer;

- the influence of low bit-rate codecs, the equipment impairment factors: subjective tests, published in ITU-T Contributions to SG 12 (SQEG), SG 15 and elsewhere;

- the expectation factor A: interpretation of information from the market place that certain systems enjoy widespread public acceptance, i.e. their "speech communication quality" is deemed to be high, although their "absolute" voice transmission quality must be considered rather low in comparison with that of the conventional telephone network. The first example is mobile communication with its phenomenal growth. The second example is double-satellite hops to places that otherwise could not be reached by telephone. (The values of the expectation factor A are provisionally chosen so that they compensate half the impairments caused by the GSM (RPE-LTP) codec respectively the long absolute delay, i.e. A = 10 respectively 20).

To see if the ETSI model in its present form gives reasonable results, some comparisons have been made with other models and certain subjective tests for the most important types of impairments. So far, the results appear promising. Further aspects of validation of the ETSI model are discussed in subclause 9.3 and in annex K.

9.2.2 Translation of R-values into GOB, POW, TME, MOS

The percentage of subscribers finding the call "good or better" (GOB), the percentage finding it "poor or worse" (POW), and the percentage enough dissatisfied to terminating it early (TME), are obtained in the ETSI model from the R-factor by means of normal error distribution functions, all having the *standard deviation* in R = 16. The *mean values* in R are 60, 45 and 36 for respectively GOB, POW, TME.

Figure 53 shows GOB(R) and POW(R) as well as the corresponding curves in CCITT Supplement 3 [43] for the BcTR model, namely results from the "Long toll", "CCITT" and "MH" investigations. As can be seen, the ETSI curves approximately represent mean values of the BcTR curves.

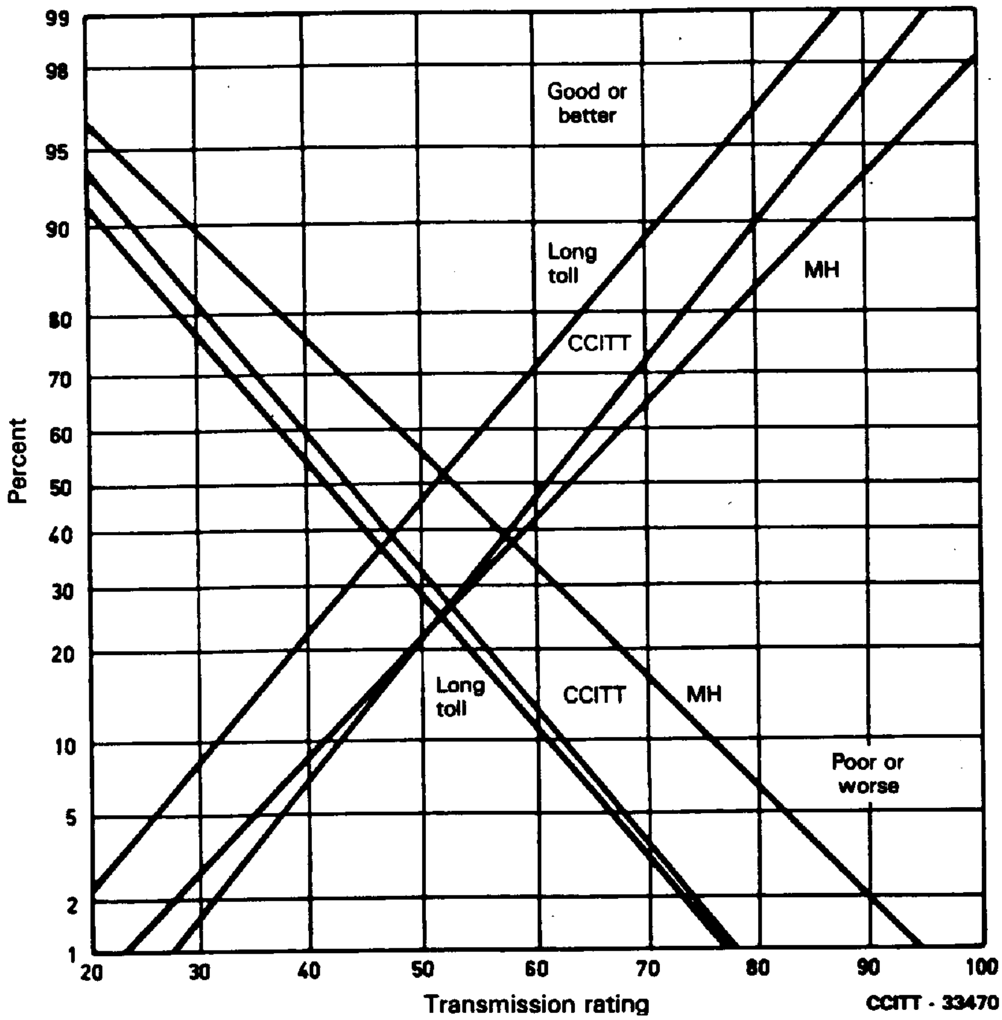
Figure 54 depicts results from "AT&T Long Toll" interviews as cited in reference [9]. The ETSI curve for "Terminating Early", TME, was chosen to be exactly the same as the one in the figure.

The BcTR model does not contain a function MOS(R). However, Annex A to ITU-T Recommendation P.11 [26] describes a very similar model, here called the TQI model, where expressions are given for GOB, POW and MOS. By means of these it was possible to derive a provisional algorithm for MOS as function of R:

For $0 < R < 100 \quad MOS = 1 + R/25 + R(R-60)(100-R) \cdot 7{,}5 \cdot 10^{-6}$
For $\quad R < 0 \quad MOS = 1;$
For $\quad R > 100 \quad MOS = 5.$

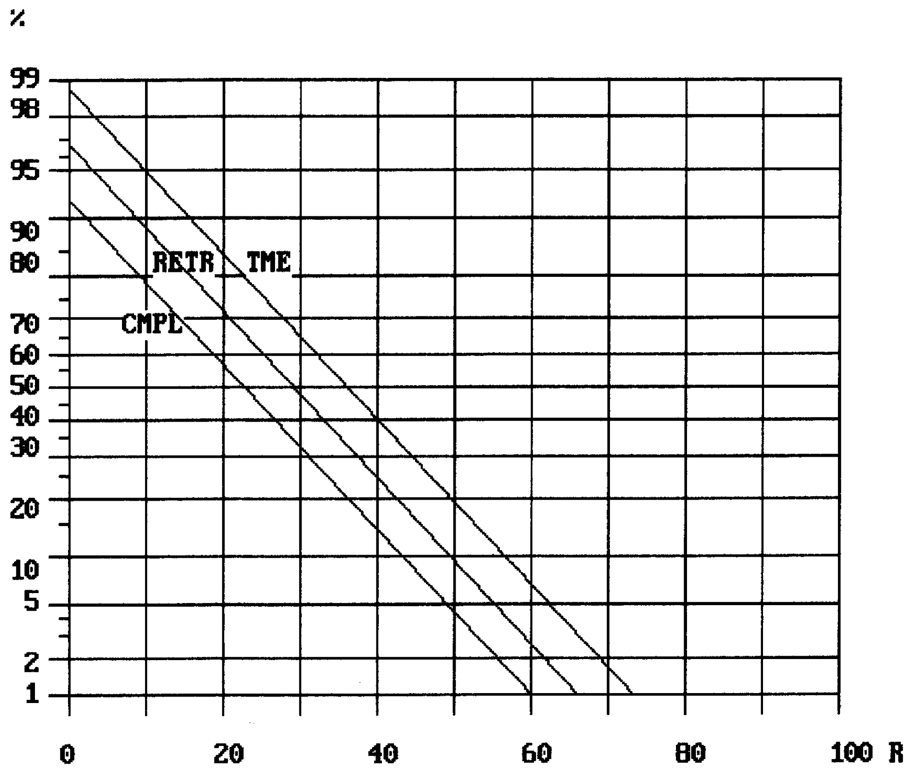
Note that according to these equations, MOS can reach a maximum value of 5. However, this is contrary to experience from actual subjective tests where a score of 5, "excellent", never is given by the participants. A more realistic maximum value appears to be MOS = 4,5. The equations above therefore were modified by a linear transformation into equ.(9.1.5 - 9.1.7). MOS(R) is depicted in figure 55.

NOTE: Subscribers' opinions vary with time and circumstances (previous experiences) so that one cannot expect a high degree of precision in actual customer surveys of GOB and POW. The same applies for subscribers' spontaneous reactions in the form of TME. Even MOS figures obtained under controlled laboratory conditions for a particular impairment may vary between one experiment and another. Therefore, the predictions for GOB, POW, TME and MOS obtained by the ETSI model should be considered as *nominal* customer reactions for guidance in transmission planning.



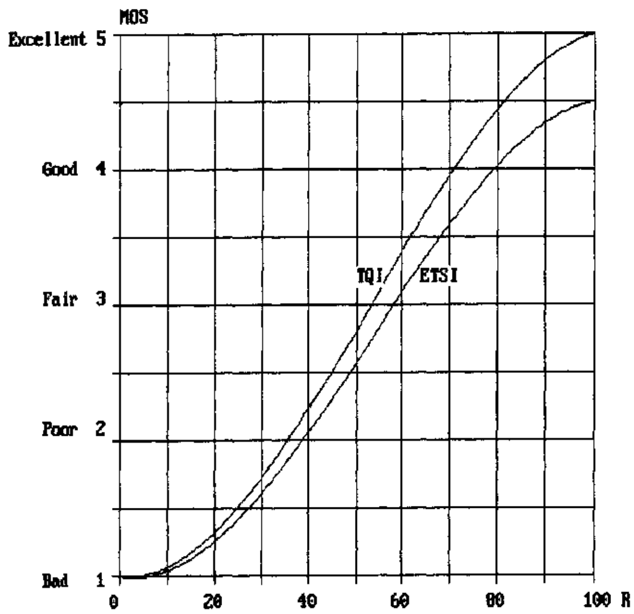
—— Bell Labs
------ Values chosen for ETSI model (as per figure 58 for example)

Figure 53: GOB(R), POW(R); Subjective tests made by Bell Labs



NOTE: TME = Terminate early; RETR = Retrial; CMPL = Complaints.

Figure 54: Customer behaviour versus the Transmission Rating factor R



NOTE: For comparison, the MOS given by the TQI model is also shown.

Figure 55: MOS as function of R in the ETSI model

9.2.3 Overall loudness rating OLR and circuit noise Nc

The comparisons are made at a point where RLR = 0 and the "noise floor" = -64 dBmp.

Figure 56 shows the *R*-factor for the BcTR model as a function of OLR and with Nc as a parameter. As can be seen, the maximum in *R* occurs around OLR = 8 - 9 dB, irrespective of the Nc value. This, however, is contrary to results from other models and from some subjective tests: when the circuit noise is high, a lower value of OLR than 8 dB is preferred. Therefore, the ETSI model has been modified accordingly, see subclause 9.1. Figure 56 depicts the result. The BcTR and the ETSI models agree quite well for OLR > 10 dB. (The ETSI model is slightly more pessimistic for higher Ncs).

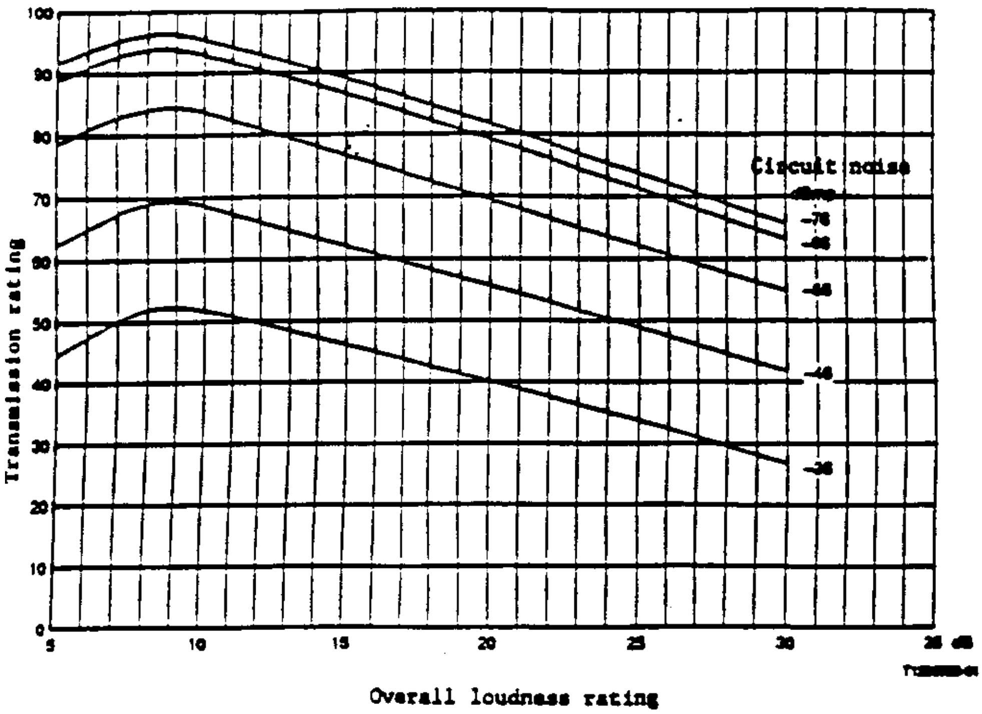


Figure 56: The *R*-factor as function of OLR and Nc at RLR = 0 dB; the BcTR model

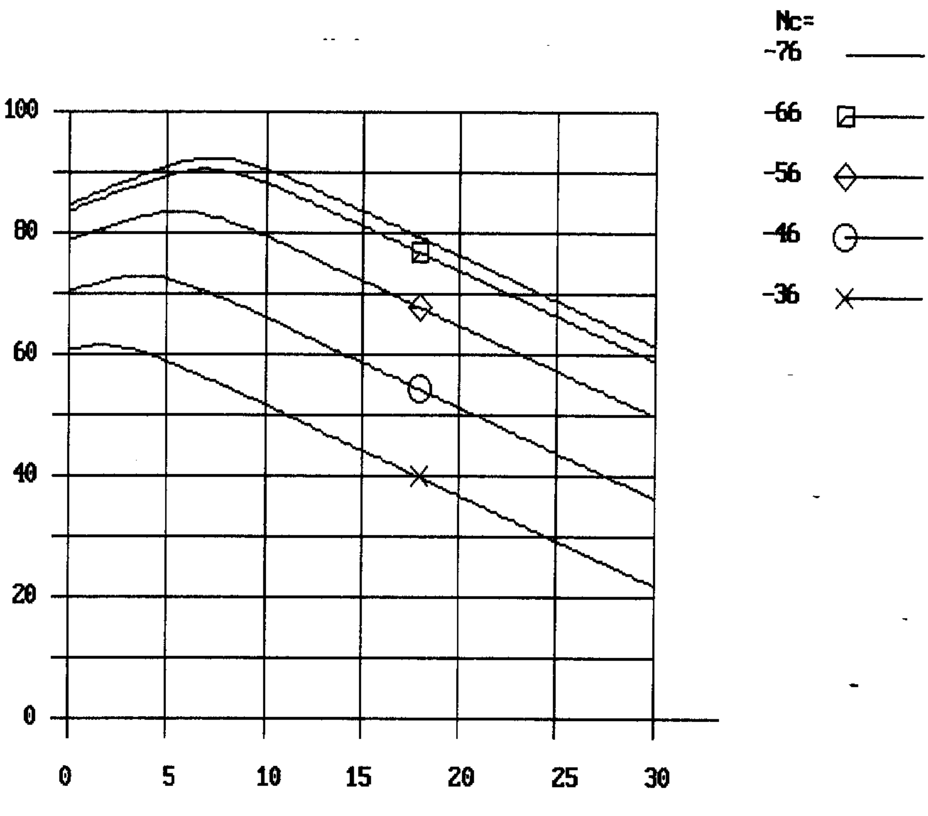
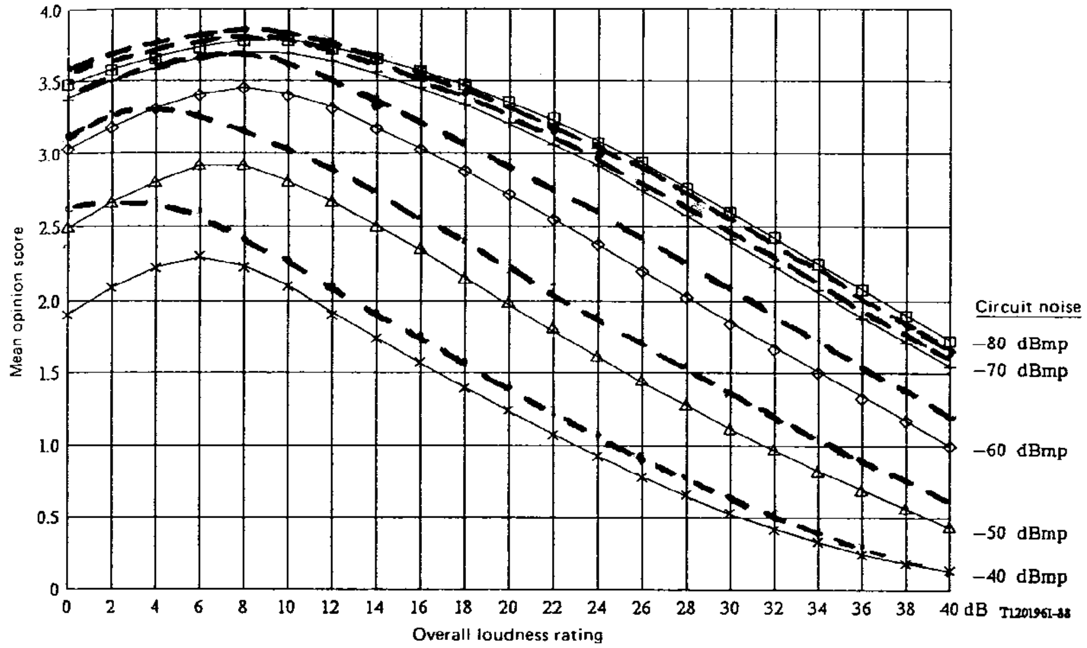


Figure 57: The *R*-factor as function of OLR and Nc at RLR = 0 dB; the ETSI model (Nc in dBmp)

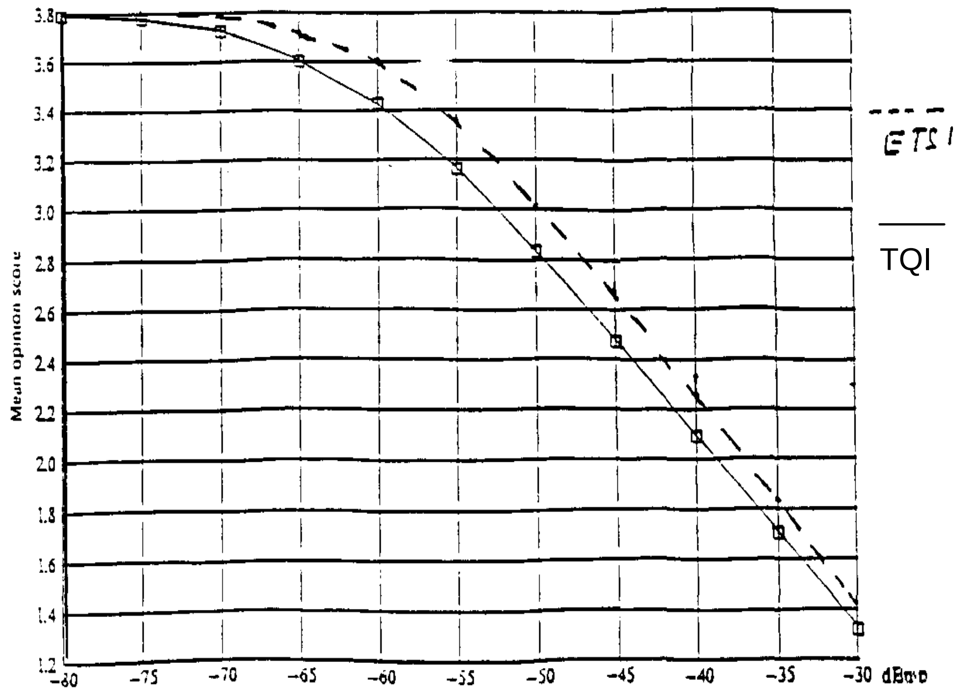
Figures 58 and 59 show a comparison between the ETSI model and the TQI model (as taken from Annex A to ITU-T Recommendation P.11 [26]) for MOS. Note that the MOS scale in Annex A is 0 - 4 instead of 1 - 5. Moreover, the TQI model assignes a MOS value to a "very good" channel which approaches 4, while the ETSI model gives MOS = 4,5 for such a channel. Therefore, the ETSI MOS values have been transposed by a linear transformation, so that 4,5 and 1 of the ETSI model correspond to respectively 4 and 0 of the TQI model.

The ETSI values are slightly more optimistic than the TQI MOS values.



- - - - - ETSI model
_____ TQI model

Figure 58: Comparison between the ETSI and the TQI models; MOS as function of OLR and Nc at RLR = 0 dB

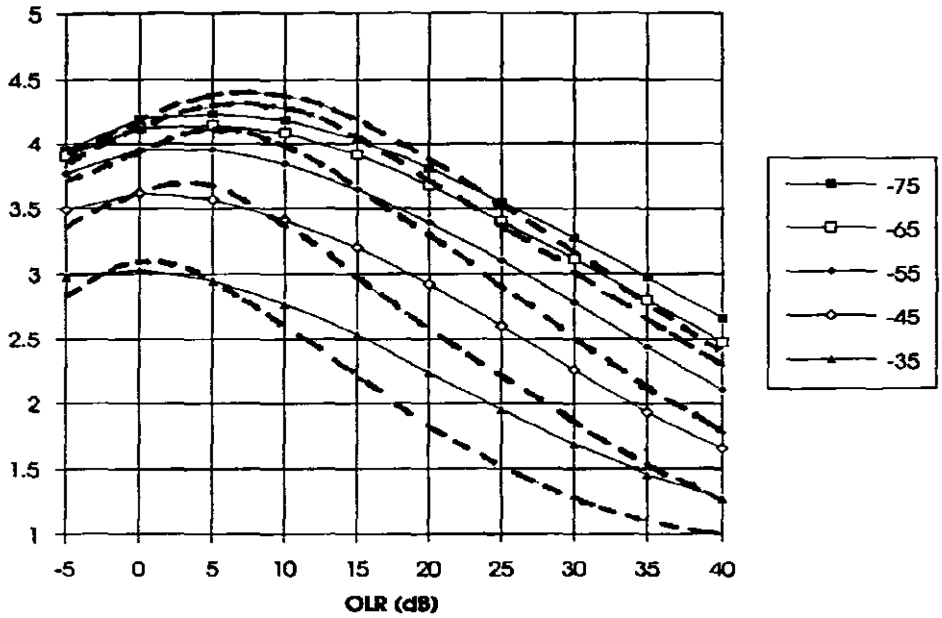


Circuit Noise

Figure 59: Comparison between the ETSI and the TQI models. MOS(Nc), OLR = 10 dB; RLR = 0 dB

Figure 60 shows a comparison between the ETSI and the CATNAP (CCITT Supplement 3 [43]) models for MOS. Here, the ETSI MOS values have been plotted directly, without any transformation.

The CATNAP model is more optimistic for very high OLR and circuit noise values. However, for normally encountered OLR and circuit noise values the agreement is fairly good.



- - - - - ETSI model
_____ CATNAP model

Figure 60: Comparison between the ETSI and CATNAP models; MOS(OLR, Nc): RLR = 0

Discussion:

The differences between the results from the ETSI model and other sources as depicted in figures 56 - 60 are not very significant and appear to be within the general uncertainty of user opinions. Moreover, in modern networks, high losses (OLR) and high circuit noise will not be a dominating impairment.

9.2.4 The influence of ambient noise via the listener sidetone

CCITT Supplement No. 11 [41] "Some effects of sidetone", presents in its figure 3 MOS values as function of LSTR for quiet connections. This is reproduced here in figure 61 together with predictions obtained from the ETSI model. (Note that the Supplement has a MOS scale 0 - 10, the model 1 - 5). The measured and predicted values agree quite well.

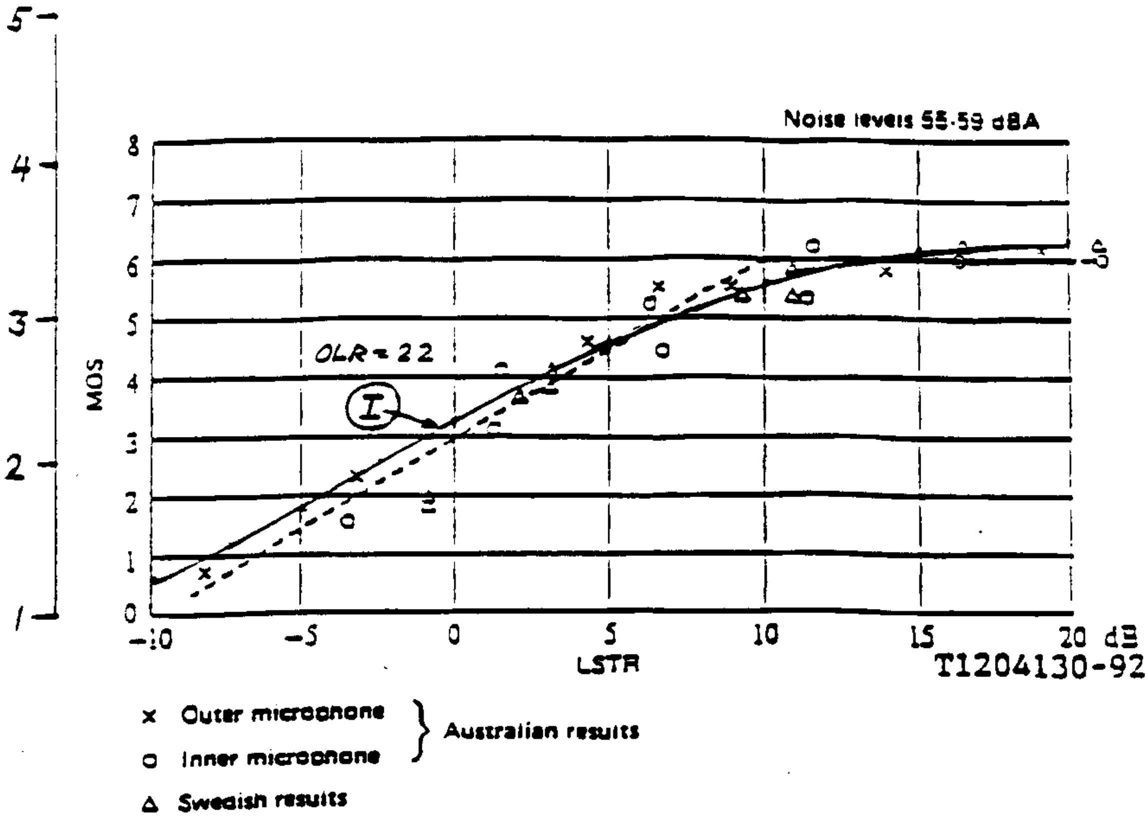


**Figure 61: MOS for listening tests as function of LSTR for a quiet connection.
Subjective tests: MOS - scale 0 - 10. (CCITT Supplement 11 [41]).
I: Prediction: MOS - scale 1 - 5. (From the ETSI model)**

9.2.5 Echoes

The impairments caused by *listener echoes, WEPL,* are characterised by exactly the same algorithms in the BcTR and the ETSI models. Therefore, it is not necessary to make a direct comparison.

For *talker echoes,* the ETSI model also uses the BcTR model, however with two minor modifications. The first was made in order to accommodate some recent subjective tests of short-time echo effects, exemplified by the comparisons in figures 62 and 63. The second applies to the fact that OLR (from the talker to the listener) cannot influence the talker's perception of his own echo. (This will be discussed in what follows).

The MOS values in figure 62 were obtained under the following test conditions as specified in reference [8]:

Low room noise, normal circuit noise and OLR = 10 dB; MOS-scale 0 - 10.

Judging from the scores given, MOS(Exp). = 8 corresponds to MOS(ETSI) = 4,5;
MOS(Exp). = 1 corresponds to MOS(ETSI) = 1.

The figure 63 test conditions as given in reference [9] have been interpreted as

> OLR = 8 dB (SLR = 6 dB, RLR = 2 dB); STMR = 12 dB and LSTR = 15 dB; Ds = Dr = 3 dB;
> Room noise = 45 dB(A) at both sides; Circuit noise Nc = 50 dBm0p.

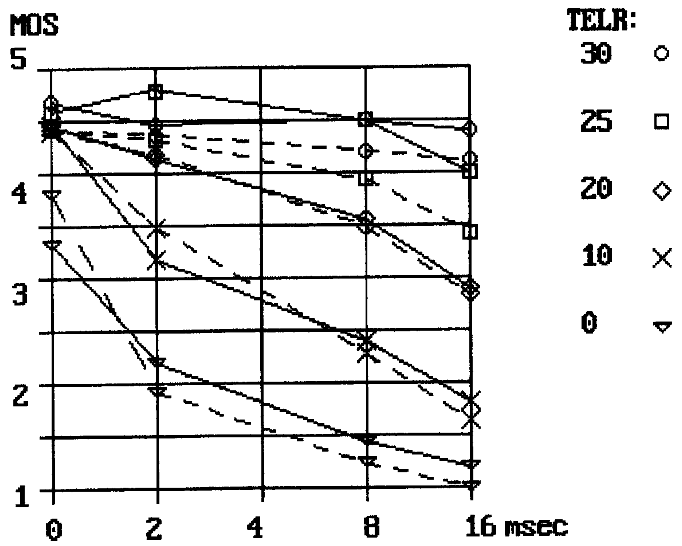
Considering the inevitable variations in subjective MOS-values, the predictions appear to be fairly good. (Note that TELR = 0 dB in figure 62 represents a very bad balance return loss which hardly occurs in actual networks unless there is a fault condition).

A particular feature of the BcTR model is that impairments caused by talker echo can be masked by circuit noise and also, to some extent, high values of Overall Loudness Rating, OLR. However, recent investigations by NTT, (see reference [57]), give a strong indication that the talker echo effects are *not* dependent on OLR! This conclusion is indeed very plausible because when a subscriber talks he may hear an echo of his own voice but not the other party who is silent. However, the echo can be more or less masked by electric circuit noise and ambient noise picked up at both send and receive side.

The impairment algorithm for talker echo in the ETSI model therefore has been adjusted in accordance with the NTT findings to be independent of OLR. However, certain "anchor points" have been used to establish a relation to previous known results of talker echo impairments. Thus, for the "nominal" case of OLR $\approx$ 10 dB the ETSI model gives almost identical values as the BcTR model. Note that the so-called 1 %-curves in CCITT Recommendation G.131 [14] for TELR limits as function of delay time T, which are based on the BcTR model for "typical" values of OLR, have proven to be very relevant for talker echo evaluation.

Figures 64 and 65 illustrate the difference in the rating *R* between the BcTR and the ETSI model for the typical delay limit T = 25 ms and circuit noise Nc = -70 dBm0p. In figure 64, OLR = 10 dB; in figure 65, OLR = 20 dB. Note that for OLR = 20 dB, the BcTR values are rather more optimistic than those from the ETSI model. (In the NTT investigation, (see reference [4]) the BcTR model also was found to give higher scores than the actual subjective test results for talker echo effects).

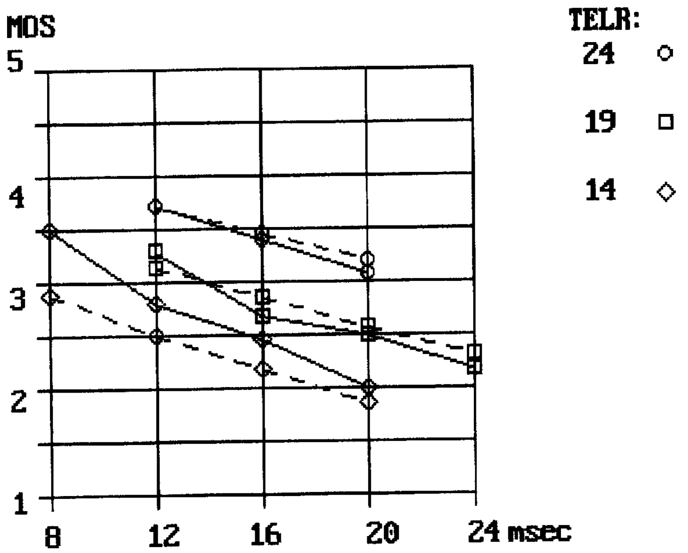
Similar curves are obtained for other values of delay and circuit noise.



_____ Subjective tests

- - - - - ETSI model

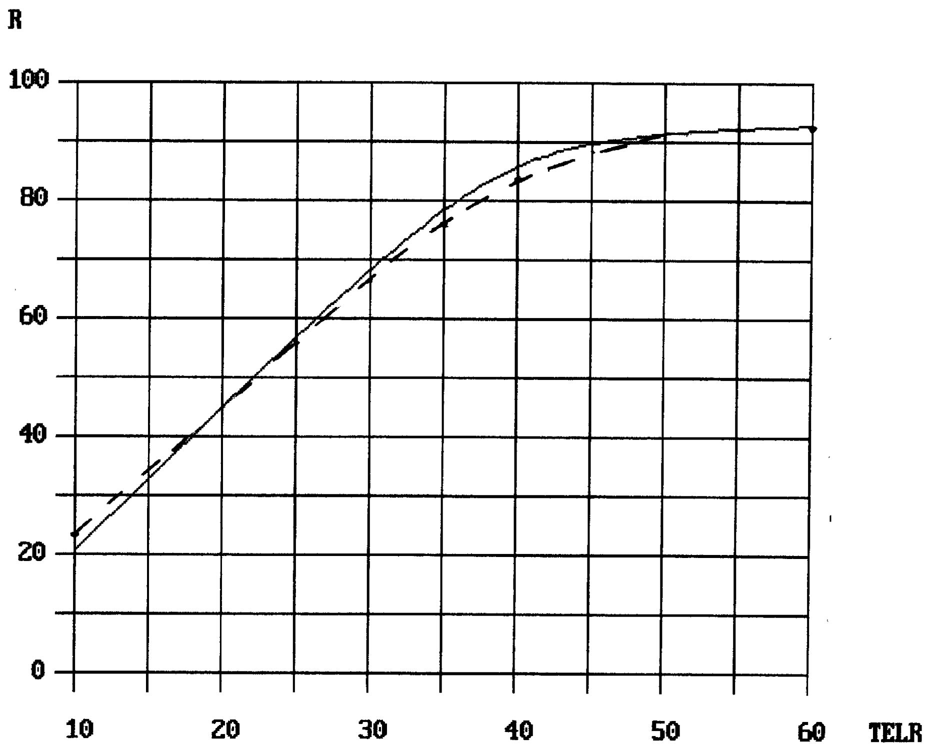
Figure 62: Subjective MOS - values, reference [4], and predicted values, ETSI model, for short-delay talker echoes. (T = mean one-way delay)



______ Subjective tests

- - - - - ETSI model

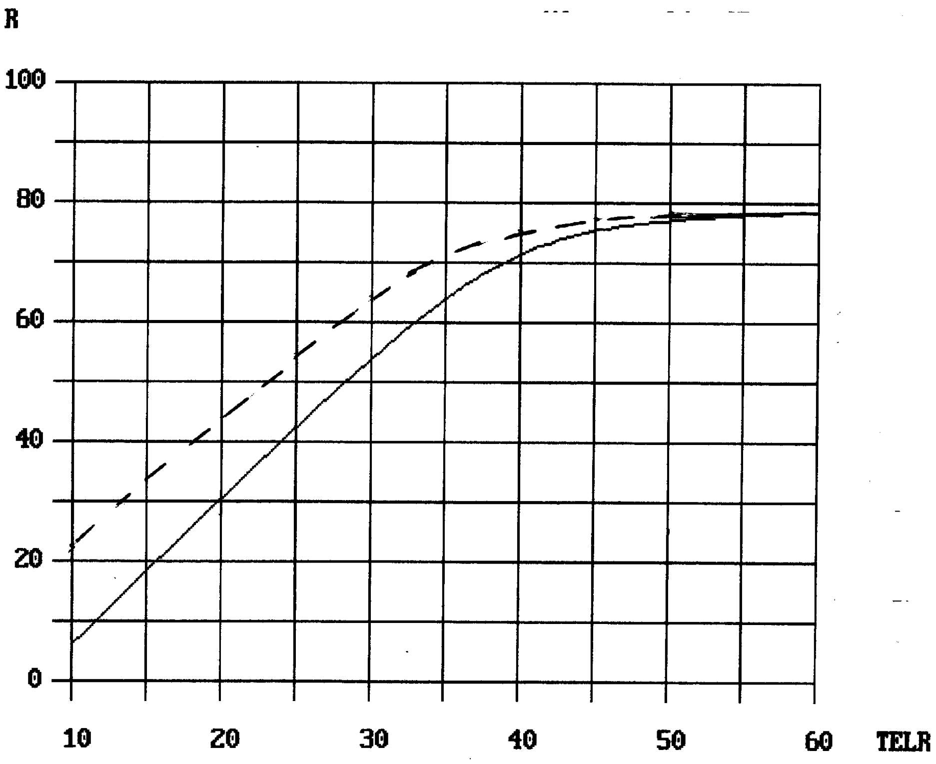
Figure 63: Subjective MOS - values, reference [57], and predicted values, ETSI model, for short-delay talker echoes. (T = mean one-way delay)



______ ETSI model

- - - - - BcTr model

Figure 64: Rating factor *R* as function of TELR. OLR = 10 dB (SLR = 7 dB, RLR = 3 dB), T = 25 ms, Nc = -70 dBm0p



_____ ETSI model

- - - - - BcTr model

Figure 65: Rating factor R as function of TELR. OLR = 20 dB (SLR = 17 dB, RLR = 3 dB), T = 25 ms, Nc = -70 dBm0p

9.2.6 Quantizing distortion unit, qdu

For quantizing distortion, caused by PCM processes, the ETSI model uses the same algorithms as the CATNAP model which gives results which are more pessimistic than those obtained from the BcTR model. Table 14 shows, as a function of qdus, the decrease dR in R-values for the BcTR model and the corresponding impairment factor Iq for the ETSI model.

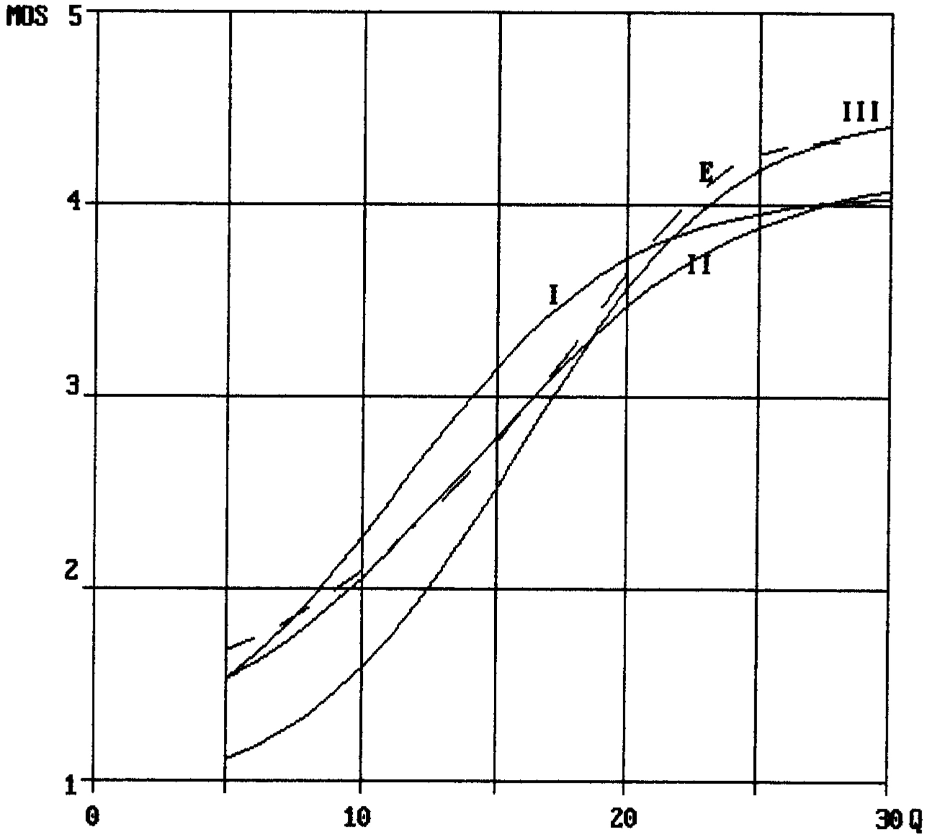
Table 14: Impairments caused by PCM quantizing distortion; qdu

dR = diminishing of the R-value, caused by qdu.
Iq = impairment factor, ETSI model.
Connection conditions: OLR = 10 dB, RLR = 0 dB, Nc = -66 dBmp.

qdu	dR	Iq
1	0	0
5	0,3	0,7
10	2,5	9,5
14	6,0	17,4
15	6,9	19,0
20	12,0	25,8
25	16,7	30,6
30	20,8	34,4
35	24,5	37,4
40	27,8	39,8

The ETSI (CATNAP) impairment factor Iq seems to be the most realistic. One reason is that the ITU-T recommended limit of qdu = 14 should represent a noticeable degradation which the BcTR model does not indicate.

As is well known, the subjective effect of quantizing distortions from PCM can be simulated by a MNRU, a Modulated Noise Reference Unit. The amount of quantizing distortion introduced by such a device is characterized by the parameter *Q* dB, and the relation between *Q* and qdu is given by equ.(9.1.24). Figure 66 depicts MOS as function of *Q*, both in the form of three different MNRU subjective test result, as well as predictions from the ETSI model. A comparison shows that the ETSI model is well within the limits of the variations between the subjective results. (The subjective test results are taken from some reference tests made in conjunction of evaluating low bit-rate codecs, see table G.3).



- - - - - E = predictions, using the ETSI computation model
——— I, II, III = results from subjective tests (described in annex G)

Figure 66: Quantizing distortion; MOS as function of *Q* dB, using a Modulated Noise Reference Unit

Note that quantizing distortion impairment is negligible for 5 PCM coding processes in tandem which is a probable limit for modern networks. Also note that low bit-rate codecs and 32 kbit/s ADPCM should not be characterised by qdus but rather with *Ie*-factors, see subclause 9.2.7 and subclause G.2, figure G.1.

9.2.7 Impairments caused by low bit-rate codecs

For obvious reasons, opinion scores for the speech quality of low bit-rate codec have a rather low degree of precision. The codec speech impairments are of many different kinds and they do not sound at all in the same way as pure circuit noise and quantizing distortion. Participants in subjective experiments do not have clear references to comparable, "quantified" impairments. It may well happen that codecs are ranked in different order with regard to speech quality in experiments made at different times and in different laboratories. Even in the simple case of pure quantizing noise a quite appreciable spread in MOS can be observed, up to 0,7 as can be seen in figure 2.1 of reference [4]. Thus, it cannot be expected that a computation model will simulate results from a particular subjective evaluation of a low bit-rate codec with any higher degree of precision. Instead, certain safety margins has to be applied when formulating rules for transmission planning.

Up to now, the data-base for evaluating impairments of low bit-rate codecs in the ETSI model has been rather limited, see reference [4]. In principle, MOS-figures for various combinations of codecs in tandem have been analysed and converted into R-values which in turn have been broken down into a basic value for the connection and impairments for the codecs. The following results have been obtained:

The ETSI model considers a low bit-rate codec of a particular type to be associated with a specific *equipment impairment factor Ie* which is additive to the other impairment factors. In annex G, comparisons have been made between subjective and predicted MOS values for various combinations of codecs. In general, the agreement is quite good, better than when the qdu method is applied.

Provisionally, the values given in table 15 apply for transmission planning. (Certain safety margins have been included). Note, however, that it is assumed that the speech levels are those for which the codecs are specified.

Table 15: Equipment impairment factors *Ie* = *K* for low bit-rate codecs in transmission planning (Provisional)

Codec	kbit/s	*K*
ADPCM	40	2
	32	7
	24	25
	16	50
LD-CELP	16	7
	12,8	20
VSELP	8	20
RPE-LTP	13	20
CELP+	6,8	25

The equipment impairment factors mentioned in table 15 have shown to be reliable enough to be used for transmission planning purposes for configurations with the codecs mentioned in this table. Annex G contains examples of the applicability of the equipment impairment factor. Correlation coefficients R > 0,92 have been observed. This is close to the ideal value R = 1 (a perfect correlation between subjective MOS and computed predicted MOS).

There are indications however that other codecs may behave differently. For such codecs the additive property assumed by the equipment impairment factor might not hold. Therefore, care should be taken in using the equipment impairment factor for other codecs in transmission planning.

9.2.8 The influence of long delays

ITU-T Recommendation G.114 [36] rules that one-way delays longer than 400 ms should be avoided in a connection as far as possible even when perfect echo cancellers are used. However, there are locations which only can be reached via long-delay connections for practical and economical reasons.

Annex B to ITU-T Recommendation G.114 [36] contains some reports of subjective investigation of the effect of long delays on subscribers´ ability to communicate with more or less difficulty. This depends of course on the communication tasks the subscribers have. However, typical subscriber reactions are depicted in figure 67 which is a reproduction of figure B-2 of ITU-T Recommendation G.114 [36]. (Investigations made by BNR on behalf of Telecom Canada).

Figure 67 also shows the prediction of POW, Poor Or Worse, percentage of customer opinions as predicted by the ETSI model. As can be seen, the model appears to be fairly realistic. (For delays longer than 800 ms the subscribers virtually have to apply "simplex" procedures instead of "duplex").

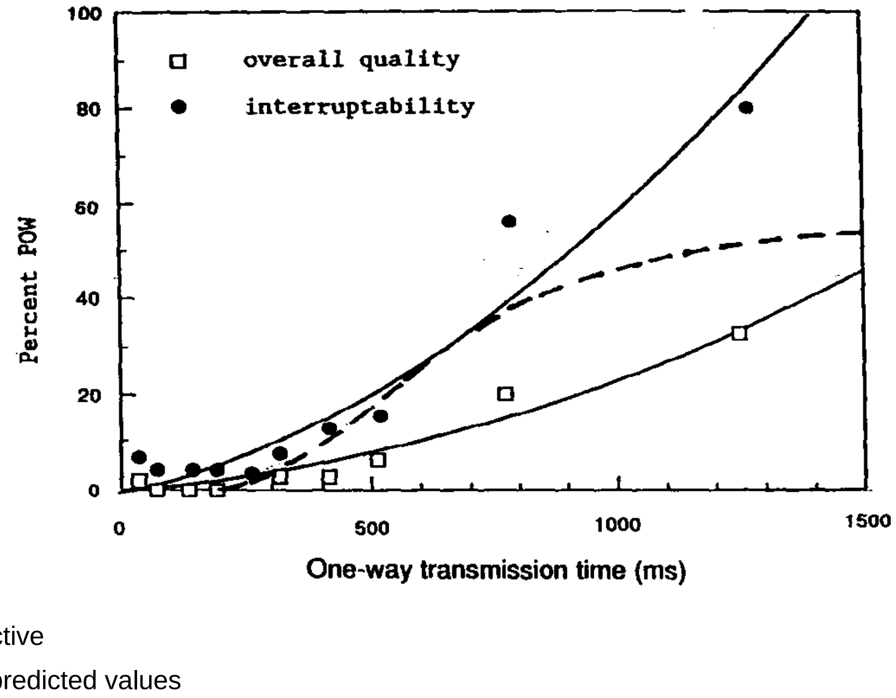


Figure 67: "Poor Or Worse" figures for long one-way delay

9.2.9 Conclusions

Comparisons have been made between results from the ETSI model and with results from other models and published subjective results for the following cases:

1) Overall Loudness Rating OLR and circuit noise Nc. (Bellcore TR model, ITU-T Recommendation P.11 [26] Annex A, TQI model, CATNAP);

2) ambient noise via listener sidetone (Subjective tests);

3) talker echo (Subjective tests for short-delay echoes, Bellcore TR model for longer-delay echoes, modified by results from subjective tests made by NTT);

4) listener echo. (Bellcore TR model);

5) quantizing distortion units (qdu). (Bellcore TR model, CATNAP);

6) low bit-rate codecs. (Subjective tests);

7) long delays. (Subjective tests, BNR in ITU-T Recommendation G.114 [36]);

8) general additivity properties (Comparisons with OPINE. Analysis of the Bellcore model with regard to OLR and circuit noise, ambient noise, quantizing distortion, talker and listener echo).

The models referred to in brackets were chosen because they are recognised in the telecommunications community. The BcTR and CATNAP models are based on extensive subjective tests. The subjective comparison tests were chosen among newly published reliable results in order to complements the investigations, especially for impairments not covered by the comparison models and in cases when the models evidently needed updating. In the future, further comparisons can of course be made with new subjective tests, for example with those on GSM half rate and ITU-T 8 kbit/s codecs.

Visual inspection of the results reveals that the differences between predictions from the ETSI model and the other models are frequently in the order of 0,2 or 0,3 MOS. An exception, a difference of up to 0,5 MOS, is for OLR in two cases: CATNAP for high noise circuits and TQI for low OLR values. The CATNAP and TQI results differ from each other in these cases.

The comparison of the ETSI model predictions with the new subjective tests shows an excellent result for low bit-rate codecs (see annex G) and a quite good one for the listener sidetone investigations. For low delay talker echo, the differences appear to be on the average 0,3 MOS, with a maximum of 0,5 MOS. For quantizing distortion the ETSI prediction is within the actual spread between different subjective tests as is shown in figure 66.

It appears feasible that the ETSI model can be used with caution in its present form as a tool in transmission planning. It is advised that the ETSI model is used in conjunction with present planning processes, rather than as a substitute. Users should be aware that verification is ongoing and that the model may change in its details as a result.

9.3 Future verification of the ETSI model

It is important, in order to achieve confidence in the model, that effort is put into a verification procedure. This should include several aspects such as:

1) formal subjective tests;

2) analysis of customer surveys;

3) a systematic collection of experiences from the application of the model in transmission planning, including comparisons with established planning rules as well as results from other computational models.

Most comparisons have so far been performed graphically. It would be highly desirable to include more numerical comparisons.

There are several reasons for a continuing verification:

- impairment types, already included in the model, may need double-checking, especially with regard to those new equipment impairment factors, which are not included in previous models, and their additivity properties in combination with other "conventional" impairments;

- it may be desirable to include additional impairment factors in the model, in particular when new types of equipment and systems are introduced in the telephone network;

- users' perception of speech communication quality as well as of voice transmission quality may change with time, necessitating some changes of parameters in the model;

- an extended verification will make planners more confident in the application of the model.

A more detailed discussion of verification procedures is given in annex K.

9.4 Application of the ETSI model in transmission planning

9.4.1 General

Establishing a transmission plan is not a stand-alone operation but must be part of the general network planning in which many considerations have to be taken. Economic and technical feasibility, a proper timetable for implementation and desirable network performance are governing factors. This means that perfectly ideal transmission conditions are not to be expected when extending and renovating old networks or even when building completely new networks. Transmission compromises have to be made. The ETSI computation model can be used during this process in several ways:

- to evaluate the combination effect of impairments in a systematic way;

- to make reasonable trade-offs between different impairments for a satisfactory performance;

- to estimate the transmission quality of particular connections, for instance those representing "typical" and "limiting" cases;

- to estimate the statistical distribution of the (traffic-weighted) speech communication quality of connections in a network as perceived by the customers;

- to estimate the effects on the transmission quality caused by various changes in a network, such as decreased subscriber line loss, increased delay, improvement of echo performance by use of a complex nominal impedance, introduction of new telephone sets, more extensive uses of echo cancellers, increased use of mobile sets, etc., etc.

- to evaluate the transmission quality in two ways:

    1) with regard to the "speech communication quality", i.e. the user' quality expectation of the particular system employed, e.g. wirebound or wireless; or
    2) with regard to the "voice transmission quality", i.e. only with direct reference to the usual quality to be expected of wirebound telephone systems, "toll quality".

(See also clause 5).

9.4.2 Interpreting subjective test results and predictions; some precautions to be observed

The ETSI computation model is fundamentally based on knowledge gained from a number of subjective tests made in the past by various organisations as well as experiences derived from actual network operations and quality surveys. One must remember that a user's perception of transmission quality depends partly on what kind of performance he (or she) has been used to, partly on what new systems, hopefully improved, he is being exposed to. Anyhow, one must be aware of the fact that customers' opinions will vary with time and circumstances.

Thus, a computational model cannot be expected to emulate each and every particular subjective test or opinion survey that is made. Rather, the opinion predictions as computed by means of the ETSI model are to be interpreted as "nominal" for reference purposes.

The ETSI model produces values of the rating factor R and predictions for the subjective quantities Mean Opinion Scores, MOS, and percentages of customers finding the quality Good Or Better, GOB, Poor Or Worse, POW, Terminating calls Early, TME.

Of these, non-experts often find GOB, POW or TME easier to interpret than MOS.

For evaluating more or less "normal" connections in transmision planning, GOB appears to be an appropriate measure, while POW and TME can be used to characterize "limiting" (i.e. from a quality point of view) connections.

In the transmission planning process itself the R-factors can be used directly. If one wishes a more detailed quality information when mobile communication links are involved, R-ratings can be stated both with and without the expectation factor A included.

NOTE: In traditional transmission planning according to ITU-T, recommended limits for individual transmission parameters are used as a guidance. Although these limits are based on experience from network operations and a number of subjective tests, subjective test values expressed as a MOS are not directly referred to in the planning rules. When the ETSI model is used as a guide it thus appears logical primarily to use limits for the R-factor in planning rather than any computed estimation of users' opinions like values of a MOS.

9.4.3 Steps in the application of the ETSI computation model

The ETSI computation algorithms are described in 9.1 in detail. To facilitate the understanding of the principles in the procedure, an overview of the steps to be taken is shown in figure 68. (In practice, the calculations will of course be made by a computer program).

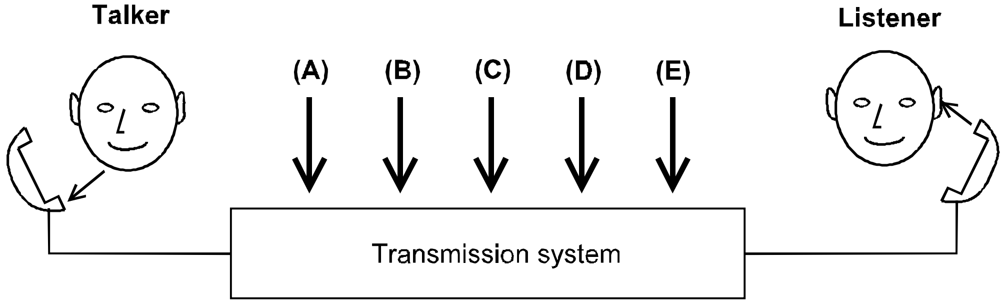


Figure 68: Steps A...E to be taken when applying the ETSI computation model

A) compute the equivalent noise and the loudness rating of the connection; compute the basic "voice-signal-to-noise ratio" *Ro*;

B) compute the "simultaneous-impairment" factor *Is* (Too high OLR, sidetone, quantizing distortion).;

C) compute the "delayed-impairment" factor *Id* (Talker and listener echo, long absolute delay).;

D) compute the "equipment-impairment" factor *Ie* (Low-bit-rate codecs etc.);

E) compute the (basic) rating factor *R = Ro-Is-Id-Ie;* add the "expectation factor" *A* (when applicable), compute GOB, POW, TME.

9.4.4 Recommended limits for the R-factor etc. to be used in transmission planning of networks

The limits in table 16 are given as a provisional guide for the transmission quality of a connection.

Table 16: Provisional guide for the speech communication quality of a connection, using the ETSI model

R	GOB %	POW %	TME %	Connection transmission quality
90	97	-	-	Very good
80	89	-	-	Good
70	73	6	-	Adequate
60	50	17	6	Limiting case
45	17	50	27	Exceptional limiting case
35	6	73	50	Customers likely to take action

To evaluate the global transmission quality of a network, it is useful to compute the traffic-weighted mean and standard deviation of the GOB for the connections. In addition, "limiting cases" should be investigated so that "hopeless" transmission situations are avoided.

When comparing different transmission cases by means of the ETSI model one must remember the basic uncertaincy of customer opinion evaluations. This means for instance that if the difference in GOB between two equipment schemes is only a few percent, this difference cannot be used as a guide to choose the optimum solution.

9.5 Statistical use of the ETSI-model for quality control of existing networks: some guidelines

9.5.1 General

The possible practical uses of the ETSI-model are numerous; the main one is to help telephone network planners to choose among various network designs or different types of equipment by applying the model to a number of reference connections, each of which includes typical parts of the network and equipment.

In all cases, those applying the ETSI model in a strictly mathematical way should not find it at all difficult to use; no theoretical knowledge is required beyond an understanding of the terms and symbols used.

On the other hand the ETSI model may prove to be less straight-forward to apply if the intention is to use the model extensively and systematically for an existing telephone network, i.e. to control the quality of service offered and to verify that the network itself is efficiently run. The reason is that in this case the parameter values have to be assigned by default and/or measured in the field, manually or automatically.

In the past, it was not practicable to use opinion models in this way due to the difficulties of covering a wide range of different connections, for which it was necessary to choose arbitrary and/or fixed values for many of the variables handled by the model. However, to put a computation model to this particular use is now beginning to be feasible and it will become even more so in the future, thanks to the development of new measuring methodologies and the establishment of databases which are capable of emulating the configuration of the network (e.g. number and location of low bit-rate codec connections).

Nevertheless, some limitations still exist:

1) the computation results from the model may differ from the result of a particular customer opinion survey, depending upon the fact that opinion surveys as well as subjective tests show a large variability. One must also be aware of possible long-term changes of speech quality perception among the network users;

2) some of the parameter values may not be readily available so they have to be based on assumptions.

It is therefore necessary to pursue the following goals (each individual network operator is responsible for selecting the most appropriate method of achieving these goals):

- to use feedback from customer surveys and analyse market reactions to voice quality factors so that the ETSI model can be up-dated, if necessary;

- to reduce the number of assumed parameters by measuring extensively as many variables as possible, giving priority to the most significant ones;

- to perform measurements which are accurate, automatic and properly planned;

- to develop methodologies for a correct use of the measurements taken and for solving problems arising from network complexity and heterogenity.

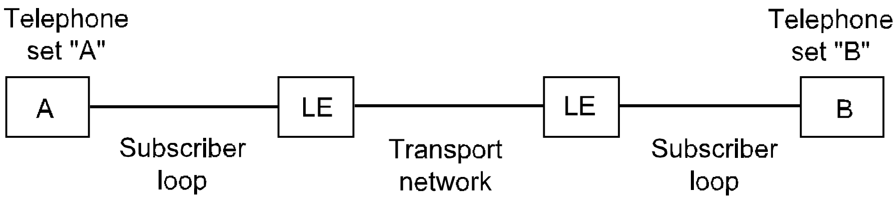
9.5.2 The statistical approach

A typical end-to-end connection (broken down into different segments) is shown in figure 69. Here, the segments depicted are the telephone sets, the subscriber loops, the local exchanges and the transport network. The latter may in turn consist of several separate segments in the form of (ITU-T) circuits.

If the transmission parameter values of the connection are known, the ETSI model can be used to for assessment of the voice transmission quality of a telephone conversation between the two subscribers. However, it is hardly feasible to make extensive measurements end-to-end. Instead, one has to rely on knowledge about the network and make certain assumptions about conditions.

Most of the parameters belong to the network and are obtained by simply adding the parameters for each segment, for instance loss, circuit noise, delay, qdu, etc.

Talker echo will depend on whether the connection is equipped with echo cancellers or not. If not, the talker echo effects perceived by a subscriber will depend on the delay but also, to some extent, on what kind of telephone set his conversation partner has, i.e. how well its input impedance is matched to the balance impedance of the local exchange. Today, the network providers have little control of this due to the liberalisation of the telecom market. This is one example of parameters that have to be based on some assumptions. Others are sidetone and room noise.



LE = Local exchange

Figure 69: Typical end-to-end telephone connection

In many cases, a transmission parameter for a connection segment is known in the form of a statistical distribution, often only specified as a mean value and a standard deviation. If a total connection parameter is obtained by adding the segment parameters, it is a straightforward procedure to obtain the mean value and standard deviation for the overall parameter, of course taking account of any mutual correlation between the segment parameters. (When many segment parameters are added, the total statistical distribution tends to approach the Gaussian (normal) distribution).

For each impairment factor *I* in the ETSI model, it should be possible to calculate the mean and standard deviation from the parameter distributions which are valid for the connections involved, at least approximately. Then the mean and standard deviation of the rating factor *R* is computed, preferably taking account of the telephone traffic weighting of the connections. As in all probability the *R*-factor has a normal distribution, its cumulative distribution can be presented and in turn be translated into percentages GOB, POW, TME.

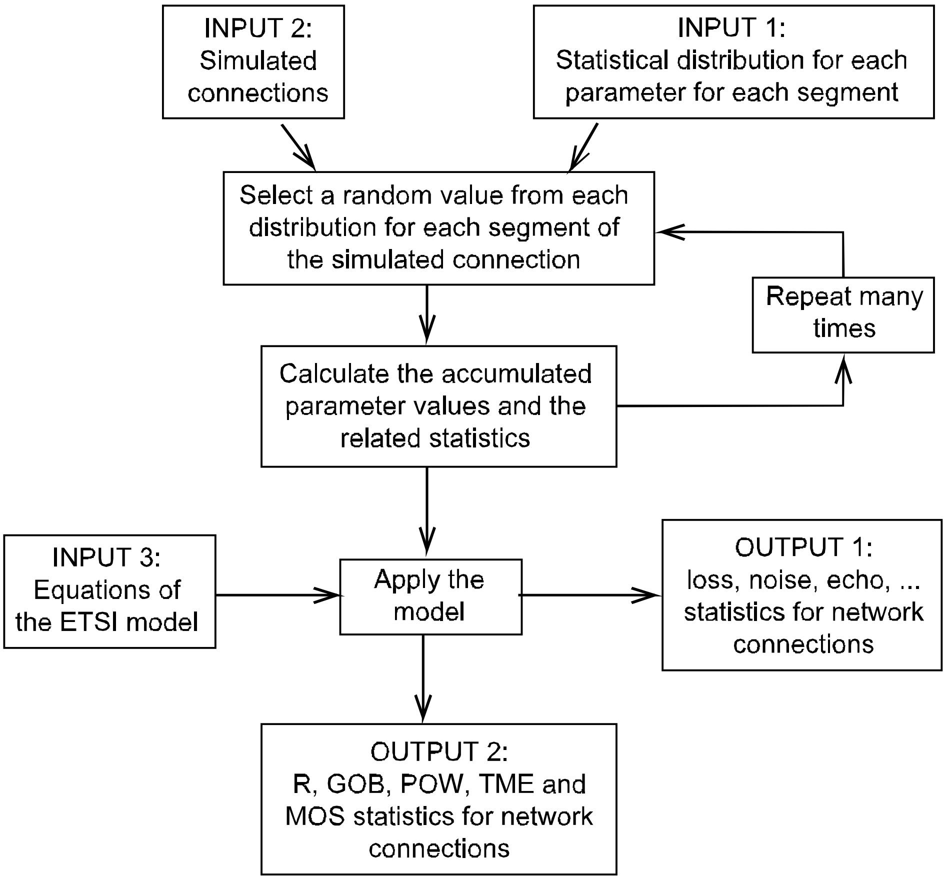
By means of modern "In-service Non-intrusive Measuring Devices" (INMD), many network transmission parameters of use for the ETSI model could be gathered, directly or indirectly. Some examples are:

- circuit noise;
- talker echo delay;
- speech levels can indirectly be translated into approximate values of SLR and RLR;
- TELR can be estimated from measurements of echo, its delay and SLR, RLR.

Much remains to be done, however.

NOTE 1: When the statistical parameter distributions have to be estimated from a limited number of measurements it is advisable to establish the confidence limits by applying a Student's t-test, as described in various handbooks on statistics.

To get a more detailed insight into the statistical transmission properties one can use a sophisticated technique for simulation such as the Monte Carlo method. Figure 70 depicts the procedure.



NOTE: The selection of "random" parameter values from each segment must be made with consideration of possible mutual correlations between some of the parameters as well as of the traffic weighting of the different connections.

Figure 70: A simulation (Monte Carlo) process for the statistical use of the ETSI model

10 Assessment of end-to-end voice transmission performance of networks

10.1 General

The most valid measure of the end-to-end voice transmission quality is to be found in the reaction of a customer as he (or she) uses the system. This can be evaluated by appropriate surveys of customer opinions. A second, complementary method is to measure some key transmission parameters in the network more or less automatically.

In reference [98] it is pointed out that automatic trouble detection and reporting are essential to the efficient operation of networks, but these purely technical measures should not be used as the exclusive indicator of quality service. To achieve quality requires more than excellent equipment, administration and maintenance. It also requires regular monitoring of changing customer values and perceptions of service.

10.2 Assessment of customer opinion

The voice transmission quality is only one aspect of the general telephone service quality. CCITT Recommendation E.125 [125] which also contains some questions dealing with the transmission quality. A more specific and detailed questionnaire regarding transmission is given in CCITT Recommendation P.82 [67].

In the questionnaires, the interviewed customer is asked to qualify the general transmission quality in four categories: "Excellent, good, fair, poor". (The reason why only four are used is that people in these kinds of field tests have found it difficult to discriminate the quality level in more detail. In controlled laboratory experiments, however, five levels can be given because then the subjects can be suitably instructed beforehand).

The customer is also asked: "Did you or the person you were talking to have any difficulty in talking or hearing over the connection?". If the answer is "yes", a more detailed questionnaire is followed so that the nature of transmission impairment(s) can be ascertained.

The aim of the ITU-T assessment Recommendations is to provide standardised procedures so that test results for international calls can be compared between Administrations. For their own use, network operators can of course design customer surveys in different ways and include more extensive investigations. Some (rather old) examples are given in references [89] and [104].

In reference [89] it is described how telephone customers were taking part in tests designed to elicit and record their evaluations of transmission quality. Results obtained in listening-only labs, conversation labs and from field tests in offices and homes were correlated with each other in order establish appropriate limits for transmission parameters for the AT&T network. To collect information in the context of actual telephone conversations, call-back interviews were made with some customers while others were asked to fill in questionnaires. Also, a special test system called "Sibyl" was employed in which Bell Lab participants, on a voluntary basis, judged various transmission impairments which were introduced in some of their normal office calls.

In customer interviews about the transmission quality of public network calls just completed it is usually not known what the actual transmission parameters were for that specific calls. Reference [104] reports a method used by AT&T, the "Measured Impairment Survey", where objective, central-office-based measurements could be compared with subjective customer perceptions of quality on the same, real telephone calls. Unlike laboratory simulations, this survey was based on actual customers using the network as they normally do. Unlike the standard attitude surveys, only a few minutes separated the customer interview and the telephone call in question. The measurement system checked customer loops (i.e. the subscriber circuit) and performed electrical and other objective measurements.

It is interesting to note that the quality evaluation programs did not solely rely on "subjecive tests under closely controlled laboratory conditions". It was found necessary to correlate these with field tests and customer experience of actual telephone network usage. This fact should perhaps be remembered today when new equipment and systems are being evaluated.

Modern networks will encompass an increasing number of connections with low bit rate codecs and mobile circuits. Customer surveys should therefore be so designed that the special transmission peculiarities of such equipment and systems also can be identified.

10.3 Objective parameter measurements

10.3.1 General

The attention of network operators is increasingly becoming focused on transmission quality, driven by a combination of customer expectations and technical developments. The trend will be towards continual assessment of quality in both public and private circuits and networks. There are two strategies which could be used to meet these requirements: The first is using **out-of-service** measurement and the second is using **in-service** monitoring**.** For out-of-service measurements, a signal is injected into one end of an idle circuit and its properties measured at the far port, allowing conclusions to be drawn about the characteristics of that circuit. For in-service monitoring, current practice involves a carefully-trained human observer monitoring brief samples of calls in the network and recording an assessment of quality.

To form a statistically accurate picture of the performance of a large network, many circuits must be measured in a short time. The cost of making frequent day-time measurements out-of-service could be prohibitive, since a circuit cannot carry revenue-earning traffic during testing. If objective measures could be mapped to customer perceptions, the need for a human expert to analyse the data would be eliminated. This would pave the way for automated in-service techniques to monitor the performance of large networks at low cost.

10.3.2 In-service non-intrusive measurement

In-service, non-intrusive measurement devices (INMDs) are utilised primarily for the measurement of voice-grade parameters such as speech level, noise level, speech echo path loss and speech echo path delay. INMDs may also be used to measure parameters associated with digital transmission systems that impact the performance of the voice-grade channels being transported.

ANSI have produced a standard for INMDs (see ANSI T1.221 [105]) suitable for certain limited applications, ITU-T Study Group 12 is currently working on a Recommendation for INMDs which will extend the work of ANSI (see ITU-T draft Recommendation P.INMD [106]). The text for this section has been taken from the ANSI and ITU-T work and the reader is referred to the ITU-T Recommendation when it becomes available for the most up-to-date text.

10.3.2.1 Application

The INMD is used as a stand-alone device or can be used as part of a network element. They may be deployed at selected switch and facility nodes in telecommunications networks to measure the in-service performance parameters of voice grade services, and to locate and analyse network anomalies. For the switched network, analysis of network anomalies is made easier when the connection information such as calling and called address digits, circuit assignments involved, etc., are known, together with the measured performance. Recording of such information does not constitute intrusion of privacy, since speech intelligence is not monitored. Other optional functions may be added to make the INMD more useful.

The INMD can only be used at a 4-wire point. In order to study conditions on the two-wire part of a subscriber line, the INMD must be connected via a 4-wire trunk on the network element (that connects to the subscriber's line under study); thus, to isolate a problem to the particular subscriber's line, some means of conveying connection information from the network element to the INMD must be employed. For facility access, the INMD monitors the MF (multi-frequency) or DTMF (dual tone, multi frequency) signalling; however, in cases such as SS7 (Signalling System No. 7), a new standard will be required to define the interface protocol. The INMD measures transmission on the path including the customer provided equipment to the point of INMD measurement access. In this way the INMD can detect transmission anomalies on the built-up connection. These anomalies can be caused by the customer environment, subscriber line, switches and trunks, including anomalies at interfaces between these network elements. In particular, the INMD can, potentially, observe anomalies that are not detectable by traditional out-of-service tests. Examples of these difficult-to-detect anomalies are:

- intermittent fading;
- acoustical feedback;
- room noise;
- defective customer equipment;
- intermittent leakage on metallic circuits;
- intermittent noise;
- design violations;
- pair gain system problems;
- digital switch level and echo control problems;
- echo problems at the line-to-trunk interface;
- tones and announcement level control problems; and
- switch translation problems that cause violations in network loss plans.

Although the INMD is potentially able to detect such anomalies, it should be noted that the INMD cannot specifically separate combined signals, e.g. room noise and cable noise, or acoustical feedback and hybrid mismatch. Similarly, the INMD cannot distinguish between a trouble on a trunk, switch, the terminating subscriber line, or the terminal equipment. Network anomalies detected by the INMD may be subsequently isolated to the affecting network element, by applicable out-of-service tests, or facility performance monitoring. Alternatively, the use of advanced signal processing techniques, such as pattern recognition, may enable the INMD to infer the source of an anomaly, e.g. by recognising particular types of noise.

As an alternative to routine testing of network elements, for the detection of network anomalies, the INMD is effective when used in call sampling mode. In this way, there is no need to maintain large routine testing data-bases. For the detection of intermittent faults, the INMD is most effective when used as a portable device. Stand-alone, facility access INMD systems, complete with proprietary interfaces to data collection and analysis software, currently exist and are being used by service providers.

10.3.2.2 Measurements

The standard measurement functions are:

- active speech level;
- noise level (psophometric weighted);
- echo loss (single or multiple reflection measurements) as defined in ITU-T Recommendation G.122 [25];
- speech echo path delay (single or multiple reflection measurements).

The above functions are for voice connections only, and do not apply to connections of short duration, where there is insufficient time to make an accurate determination of the values, or where a connection is not completed.

Additional functions being studied are:

1) originating and terminating address and carrier identification digits (switched connections only);
2) facility or circuit identification;
3) time and duration of connection;
4) signal classification (voice/data/other);
5) customer identification (dedicated circuits only);
6) DS1 performance measurements;
7) 3 kHz flat noise level;
8) connection disposition measurements;
9) speech activity factor;
10) data analysis and reports;
11) saturation clipping;
12) speech echo path loss;
13) echo path loss;
14) measurement interval.

Additional functions may arise as this area is studied further.

10.3.2.3 Testing

The testing of the accuracy and reliability of these measurements under a wide range of realistic conditions is essential to the successful deployment of INMDs. It is likely that that the ITU-T Recommendation will split the requirements for testing into four classes of INMD; for operation on national, international and non-linear and/or time-invariant networks.

Class A: Limited to measure short delay routes containing analogue and 64 kbit/s PCM components only (i.e. no low bit-rate codecs) and no echo control devices;

Class B: Limited to measure moderate delay networks (< 150 ms) that include echo control devices;

Class C: For use in international/global networks that may include signal processing devices such as echo control devices and speech compression devices (e.g. DCME and ADPCM) which incorporate waveform encoding but do not include hybrid coders (e.g. LPC) and vocoders;

Class D: For use on arbitrary possibly non-linear & time-invariant, networks which include processing devices such as LPC coders.

NOTE: INMDs designed for operation on international networks and networks that are non-linear and/or time invariant are likely to be much more complicated than those not so designed. It is felt that these devices need to meet separate more stringent requirements whilst not impeding the development of devices for national networks.

10.3.2.4 Modelling

The output from an INMD can be used as the inputs for a transmission planning model. However some parameters that affect the transmission performance are not explicitly available at the point of measurement. These "indirect measures" must be estimated from the measurements made by INMD.

The objective measurements will need conversion into end-to-end parameters suitable for a transmission planning model e.g.

- circuit noise;
- talker echo delay;
- speech levels can indirectly be translated into approximate values of SLR;
- RLR needs to be estimated from knowledge of the network;
- TELR can be estimated from SLR, RLR and measurements of echo loss.

Further work is required to estimate all the required end-to-end parameters from the INMD measurements.

10.3.3 Speech Transmission Index (STI)

The Speech Transmission Index (STI) is an objective out-of-service measure which can be used for speech transmission quality assessment, see references [78] and [80]. The measure is known to have a high correlation with subjectively measured speech intelligibility. The STI method can be used to assess the effect on speech transmission quality with regard to intelligibility over a wide range of impairments (frequency bandwidth, echo, reverberation, room noise, circult noise, linear and nonlinear distortion etc.).

The STI method uses an artificial speech slgnal. This STI-signal is a composite signal consisting of octave bandfiltered noise carriers (centre frequencies 125, 250, 500, 1 000, 2 000, 4 000 and 8 000 Hz) that are amplitude modulated with a sinusoid. The long-term spectrum of the artificial speech signal resembles the long-term spectrum of natural speech. The amplitude modulation of the noise carriers simulates the modulatlons (envelopes) that can be found in natural speech. If the circuit under test has impairments that affect the envelope of the signal, the speech intelligibility, an important aspect of the speech transmission quality, will be degraded.

In a STI-measurement the artificial speech signal is injected into the near end of the circuit under test and the modulation depth for the noise carriers after transmission is measured at the far end. The ratio between the modulation depth in the original signal and the modulation depth of the received signal (the so called Modulation Transfer Function (MTF) is calculated for all noise octaves. By weighting of the MTFs and accounting for threshold effects and masking effects between adjacent octaves, the STI is produced. This is a value in the range 0 (bad) to 1 (excellent), which is an objective measure of the parameter "intelligibility" in the speech transmission quality.

Annex A: Impedance strategy in 2-wire networks

A.1 Introduction

In order to provide a network with good echo and sidetone performance it is necessary to employ a certain impedance strategy in the two-wire analogue parts of the network where the signals are transmitted simultaneously in both directions on the same pair of wires. It is of course necessary to separate the signals (transmit and receive) from each other at the ends of a two-wire section and this is done by means of so-called hybrids which have in principle four ports:

- the 2-wire port (2);
- the 4-wire transmit port (T);
- the 4-wire receive port (R);
- the balance network port.

At an analogue subscriber line connected to a digital exchange, the hybrids are situated in the telephone set and in the exchange line circuit.

The separation between transmit and receive signals by means of a hybrid depends mainly on the impedance matching at the 2-wire terminal of the hybrid between 1) the impedance Z connected to the 2-wire port of the hybrid and 2) the so-called balance impedance Zb of the hybrid. The transmission properties of the hybrid are specified by the following parameters:

- Input impedance at the 2-wire port (2): Zo
- Balance impedance: Zb
- Matched loss between ports (2) and (R): Lr
- Matched loss between ports (T) and (2): Lt

"Matched loss" implies that the loss is to be measured for the special case when the impedance Z is made equal to Zo.

The loss Ltr between ports (T) and (R) for the case when Z is not equal to Zb is given by the relation:

$$Ltr = Lt + Lr + Lbr \qquad \text{(A.1)}$$

where

$$Lbr = 20 \cdot \lg\left|\frac{Zo + Zb}{2Zo} \cdot \frac{Z + Zo}{Z - Zb}\right| \ \text{dB} \qquad \text{(A.2)}$$

NOTE: Lbr is in general a function of frequency. Equs.(A.1) and (A.2) apply for all possible designs of hybrids.

Very often the impedances Zo and Zb are fairly close to each other so that Lbr can be replaced by the return loss

$$Lbr \approx Lr = 20 \cdot \lg\left|\frac{Z + Zb}{Z - Zb}\right| \qquad \text{(A.3)}$$

At the digital exchange line circuit, ports (T) and (R) correspond to the digital 4-wire ports. The impedance Z consists of the input impedance of the subscriber line, terminated by the input impedance of the telephone set. The loss Ltr is the talker echo signal attenuation.

At the telephone set, port (T) corresponds to the microphone, port (R) to the earphone receiver. The impedance Z is the input impedance of the subscriber cable, terminated by the input impedance of the digital exchange. The loss Ltr is the sidetone attenuation.

Clause A.2 describes how the talker echo and sidetone loudness rating are derived from the weighted averages of the respective Ltr values.

It is immediately apparent that the impedance *Z* must not vary too much from the balance impedance *Zb* if the balance loss is to be kept high. To achieve this with different cables types of varying lengths the terminating impedances have to be chosen with care. This will be discussed in subclauses A.3 and A.4.

A.2 Calculation of TELR, STMR and LSTR

The Talker Echo Loudness Rating TELR is calculated according to the formula

$$TELR = SLR + RLR + Le \qquad \text{(A.4)}$$

where SLR, RLR are the Send and Receive Loudness Ratings from the talker's side, referred to the 0 dBr digital transmit and receive ports of the exchange, and *Le* is the weighted average of *Ltr* (see subclause 6.5.2).

The Sidetone Masking Rating STMR can be computed for analogue terminals in a similar way, according to Annex A of ITU-T Recommendation G.111 [34]

$$STMR = SLR\ (set) + RLR(set) - 1 + Lst \qquad \text{(A.5)}$$

where SLR(set) and RLR(set) refer to the telephone set itself at the subscriber cable, and *Lst* is the weighted average of *Ltr.*

The Listener Sidetone Rating LSTR is

$$LSTR = STMR + D \qquad \text{(A.6)}$$

where *D* is the weighted average of DELSM (difference in sensitivities between direct and diffuse sound of the handset microphone, see subclause 6.4.3).

See subclause 6.4 for more information.

In ITU-T Recommendation G.122 [25], the weighting formula for TELR is specified for a linear frequency scale. However, *Le* and *Lst* can also be computed as averages over a *logarithmic* frequency scale, preferably using a third-octave division, giving 14 points between 200 Hz and 4 000 Hz. The general equation is

$$L = -\frac{10}{m}.\lg\left\{\sum_{i=1}^{14} Ki.10^{-0,1mLtr}\right\} \qquad \text{(A.7)}$$

The coefficients *Ki* are given in table A.1.

For calculation of *Le* (talker echo), $m = 1$; for calculation of *Lst* (sidetone), $m = 0,2$.

Table A.1: Weighting coefficients *Ki* for talker echo (*Le)* and sidetone (*Lst*)

i	Fi (Hz)	*Ki*(*Le*)	*Ki*(*Lst*)
1	200	0	0
2	250	0	0,01
3	315	0,05	0,02
4	400	0,1	0,03
5	500	0,1	0,04
6	630	0,1	0,05
7	800	0,1	0,08
8	1 000	0,1	0,12
9	1 250	0,1	0,12
10	1 600	0,1	0,12
11	2 000	0,1	0,12
12	2 500	0,1	0,12
13	3 150	0,05	0,12
14	4 000	0	0,05

Note the difference in the weighting between echo and sidetone. The talker echo weighting is flat, while the sidetone weighting has an emphasis on the frequencies above 800 Hz. This implies that for talker echo, an impedance matching is needed over the total speech band, while for sidetone only the range above 800 Hz is of importance.

A.3 Formulas for unloaded cables

For the sake of completeness, the impedance formulas for cables will be presented.

The unloaded subscriber cable parameters are:

r W/km;
c F/km;
l H/km (*l* is in the order of 0,5 mH/km and can often be neglected);
L km length.

The image (characteristic) impedance of the cable is

$$Zc = \sqrt{\frac{r + j\omega l}{j\omega c}} \qquad \text{(A.8)}$$

where $\omega = 2\pi f$; f in Hz.

The image attenuation of the cable is

$$A = L \cdot \mathrm{Re}\left\{\sqrt{(r + j\omega l)\, j\omega c}\right\} \quad \text{nepers} \qquad \text{(A.9)}$$

The image phase shift of the cable is

$$B = L \cdot \mathrm{Im}\left\{\sqrt{(r + j\omega l)\, j\omega c}\right\} \quad \text{radians} \qquad \text{(A.10)}$$

If the cable is terminated by an impedance exactly equal to *Zc*, the cable input impedance will also be equal to *Zc*. If the cable is terminated by an impedance Z_T, its input impedance will be

$$Z = \frac{Z_T + Z_c \cdot tgh(A + jB)}{1 + \dfrac{Z_T}{Z_c} \cdot tgh(A + jB)} \qquad \text{(A.11)}$$

A.4 Examples of cables terminated by different impedances; a compromise nominal impedance

The problem is to choose the input impedance of the digital local exchange in such a way that the sidetone performance of the telephone set becomes adequate, and to choose the input impedance of the telephone set so that the talker echo for the subscriber at the other end is adequately suppressed.

It can be seen from equ.(A.8) that the characteristic impedance of the unloaded cable is inversely proportional to the cable diameter and inversely proportional to the square root of the cable capacitance. Therefore, it is not possible to find *one* unique nominal impedance to match exactly all the types of cables that usually are to be found in the network of a service provider. Moreover, many analogue networks contain old equipment with an impedance of nominally 600 Ω resistive, in practice often slightly inductive. These impedances must also be considered when a compromise nominal impedance is determined. Reference [91] gives an overview of the problems and discusses the methodology. Here, some impedance curves will be presented as an illustration for a typical unloaded subscriber cable with the data: 0,5 mm copper, 50 nF/km.

The cable terminating impedances tested include first of all 600 Ω resistive but also examples of nominal complex impedances proposed in reference [91]:

a) 275 Ω in series with a parallel combination of 1 200 Ω and 200 nF.

   This is a good compromise when the balance impedance can be chosen without regard to 600 Ω equipment;

b) 275 Ω in series with a parallel combination of 820 Ω and 150 nF.

   This impedance provides a better overall balance when one also have to consider old, slightly inductive telephone sets which do not have a better return loss than 14 dB against 600 Ω.

Telephone sets of old design may have rather peculiar input impedances. For the case of illustration, however, the following simplified impedances have been chosen as examples:

1) 600 Ω resistive, no inductive component;
2) 600 Ω in parallel with 1 200 mH;
3) 200 Ω in series with a parallel combination of 300 Ω and 500 mH;
4) 400 Ω in series with a parallel combination of 500 Ω and 62 mH.

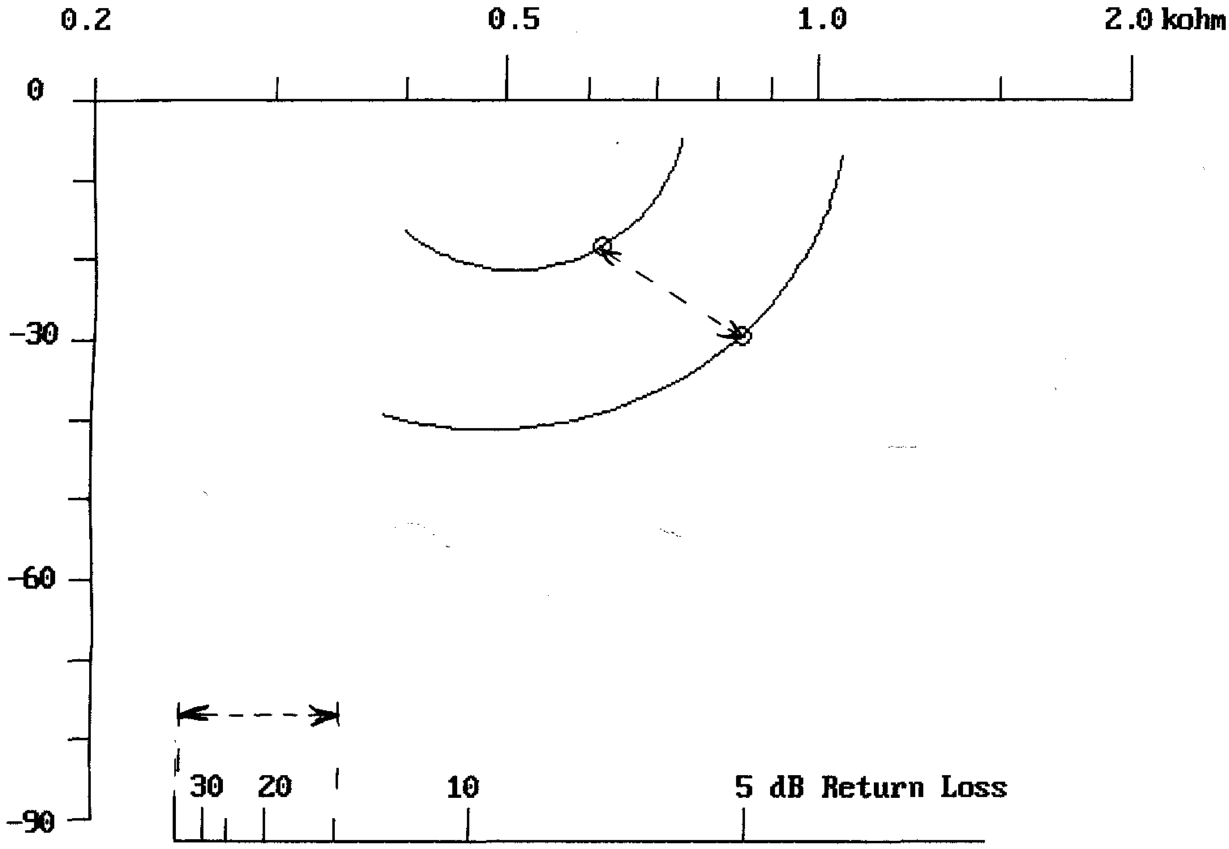
Cases 2), 3) and 4) all have a return loss > 14 dB against 600 Ω for the range 200 Hz to 4 000 Hz.

NOTE: The sidetone balance impedance (Zso) usually is more well-behaved, like a capacitive-complex, three element network.

The impedance diagrams are drawn in a special, polar form, with the abscissa and ordinate

$$X = k \cdot \ln|Z| \; ; \quad Y = k \cdot \arg\{Z\} \text{ radians}; \quad k \text{ is a constant} \qquad \text{(A.12)}$$

Note that in this type of diagrams the distance between two impedance points is, with good accuracy, a direct measure of the return loss. The return loss can be read from the "return loss" scale as shown in figure A.1.



NOTE: X-axis = Absolute value in kΩ; Y-axis argument in degrees of impedance.

Figure A.1: "Polar" impedance diagram where the return loss between two impedance points can be read from the "return loss distance scale". (Shown here 15 dB return loss)

Figure A.2 shows the cable input impedance for a 600 Ω termination for cable length 0 - 4 km and similarly, figures A.3 and A.4 with terminations a) and b) respectively. Figure A.5 depicts these impedances together with the examples of inductive sets 1) - 4).

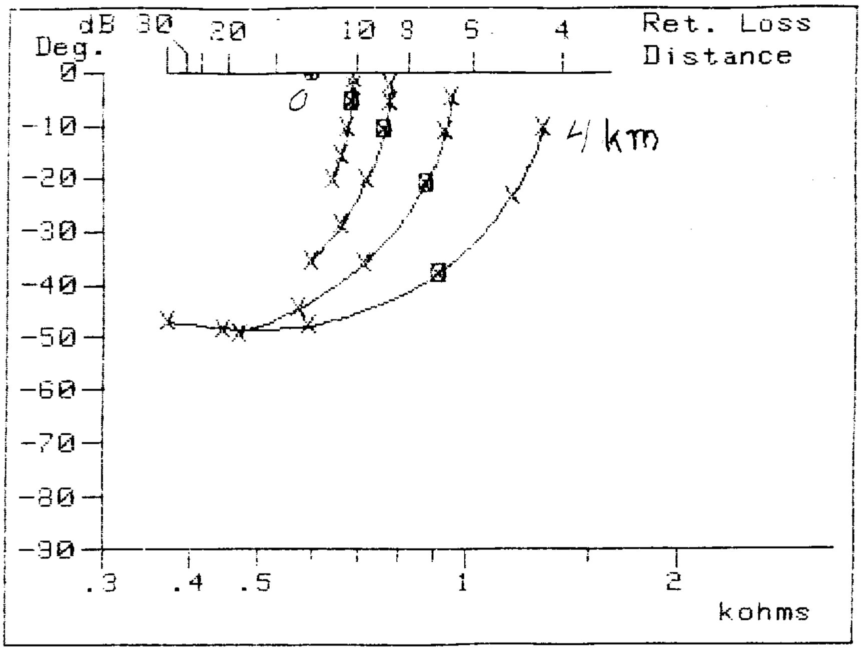


Figure A.2: Input impedance of cable terminated by 600 Ω resistive load. Length 0, 0,5, 1, 2 and 4 km. 0,5 mm Cu, 50 nF/km. Frequency markings at 0,2, 0,5, 1, 2, 3, 4 kHz

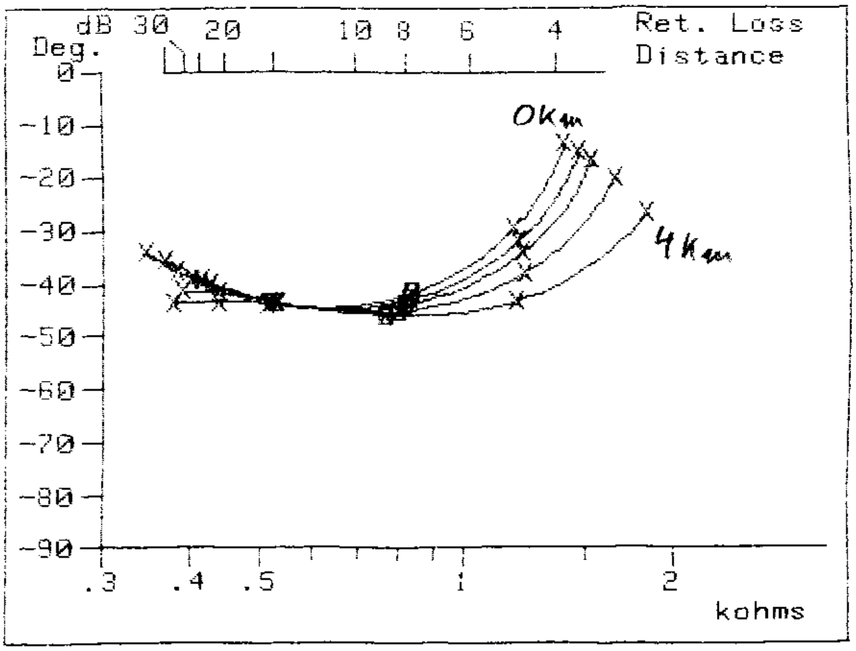


Figure A.3: Input impedance of cable terminated by impedance a). Length 0, 0,5, 1, 2 and 4 km. 0,5 mm Cu, 50 nF/km. Frequency markings at 0,2, 0,5, 1, 2, 3, 4 kHz

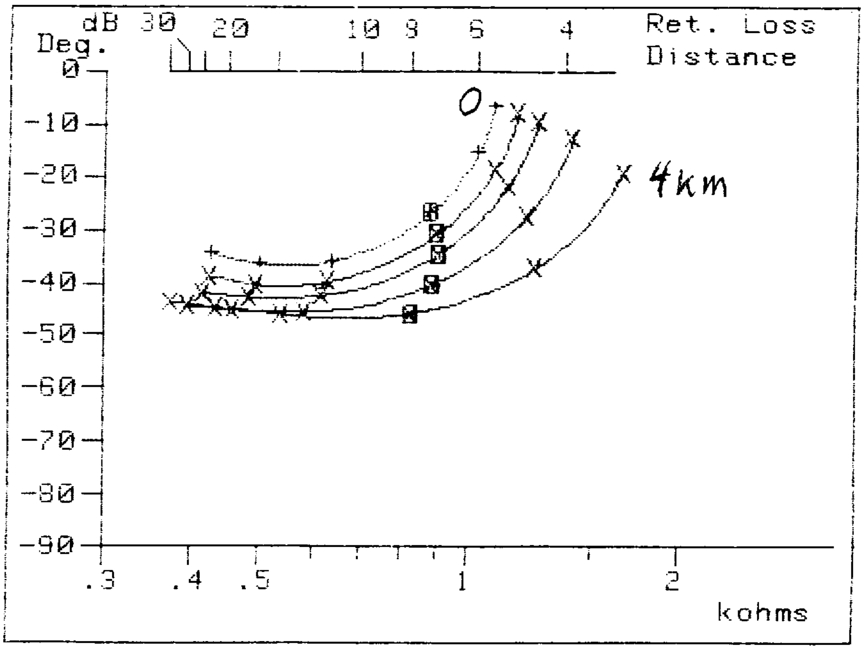


Figure A.4: Input impedance of cable terminated by impedance b). Length 0, 0,5, 1, 2 and 4 km. 0,5 mm Cu, 50 nF/km. Frequency markings at 0,2, 0,5, 1, 2, 3, 4 kHz

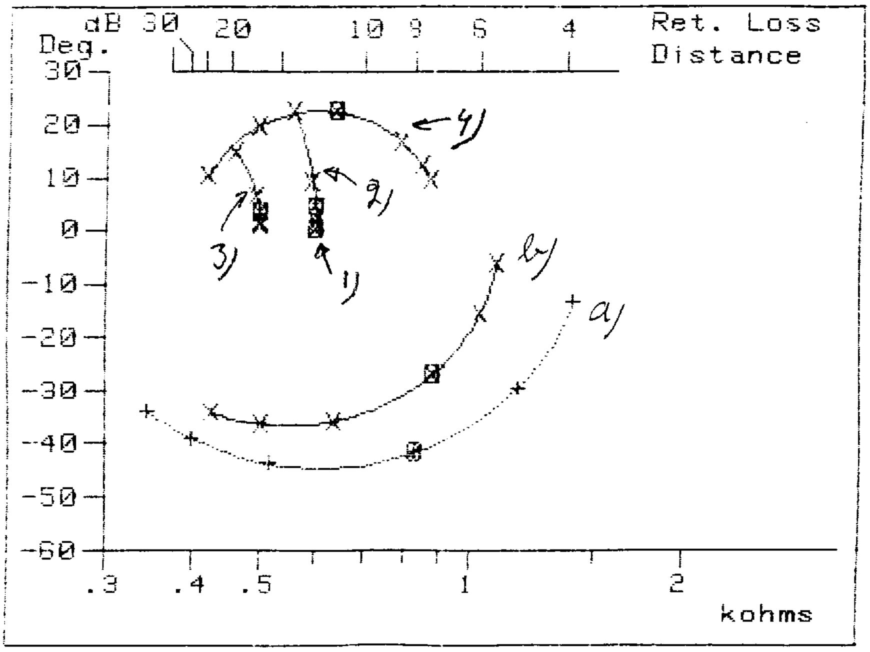


Figure A.5: Impedances of telephone sets 1) to 4) and nominal impedances a) and b). Frequency markings at 0,2, 0,5, 1, 2, 3, 4 kHz

A.5 Conclusions regarding choice of nominal impedances

Figure A.2 shows that a 600 Ω termination results in a rather wide spread of the cable input impedances when the cable length varies. From figures A.3, A.4 and A.5 it can be seen that for varying cable lengths, impedance a) would be best if old equipment could be disregarded. However, as this is not the case, impedance b) appears to be a good compromise just as is proposed in reference [91]. Another observation is that the choice of nominal impedance is not very critical. In Europe, a harmonised impedance with the value 270 Ω in series with a parallel combination of 750 Ω and 150 nF has been agreed within ETSI.

A.6 A note on the difference between the actual and the nominal input impedance of a digital exchange

In general, the actual (2-wire) input impedance of of a digital exchange is not exactly equal to its nominal (half-channel) input impedance. This is due to reflections at the far-end of the exchange. To illustrate these phenomena, some calculations have been made on a digital PBX configuration, depicted in figure A.6. The PBX is connected 2-wire to a local digital exchange (LEX).

Figure A.7 shows the input impedance *Z1* if there would be a perfect balancing at the exchange port of the PBX, a case that is hardly ever obtained in practice. Figure A.8 depicts the case with a not quite perfect balance matching between the PBX and the local exchange and its line. (Realistic PBX and LEX data have been used but it would take too much space to specify them here).

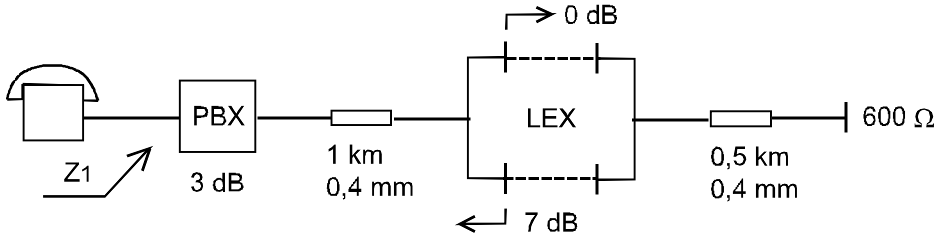


Figure A.6: Example of a digital PBX connected two-wire to a digital local exchange (LEX)

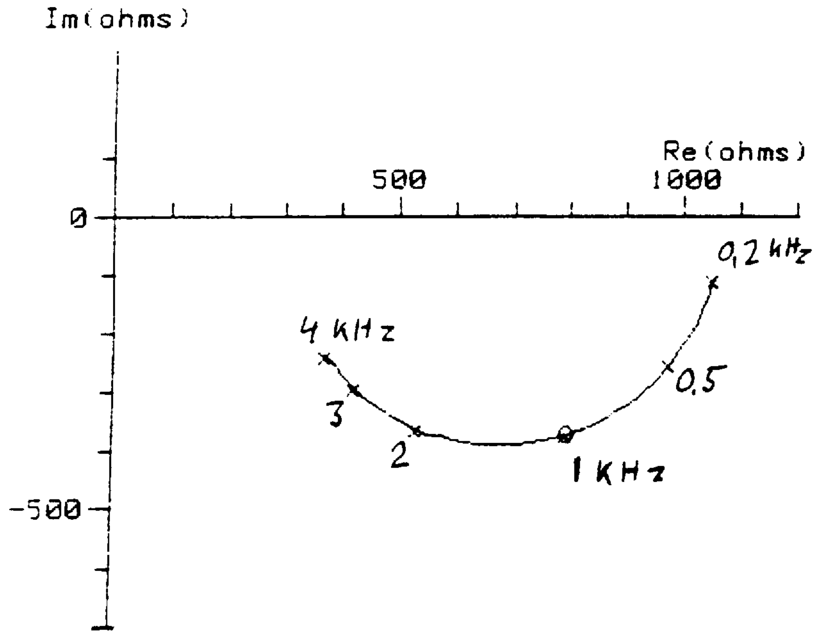
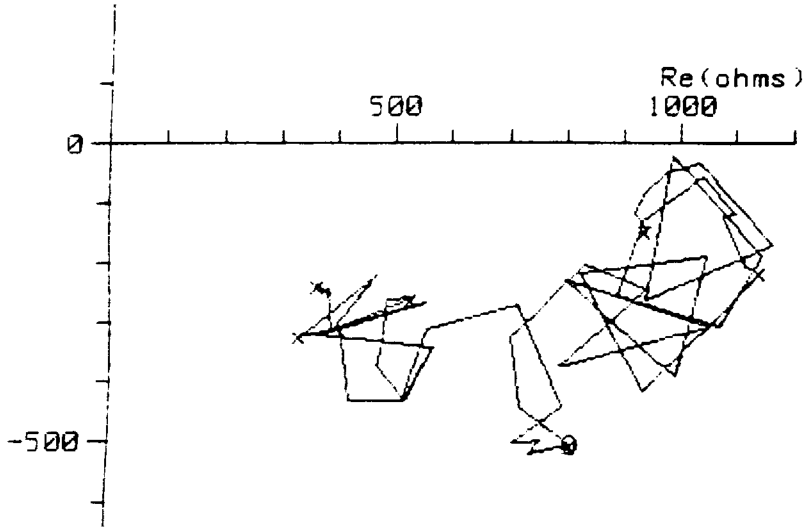


Figure A.7: Input impedance *Z1* seen by the telephone in figure A.6; perfect balancing at the exchange port of the PBX

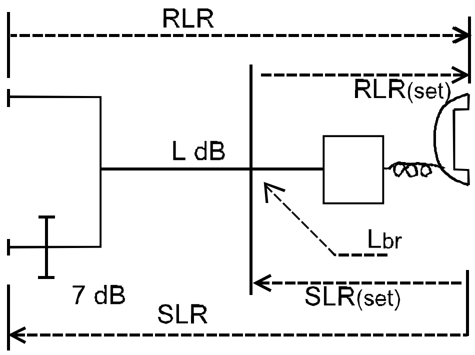


NOTE: The diagram in figure A.8 is based on impedance calculations at 54 frequencies, spaced 1/12 octave apart, in the range 200 Hz to 4 000 Hz. For simplicity, the impedance points in the diagram have been connected by straight lines. In reality, due to the rapid change of phase with frequency in the exchange 4-wire loop, the impedance curve between two points will consist of arcs-segments of circles.

Figure A.8: Input impedance *Z1* seen by the telephone in figure A.6; not quite perfect balancing at the exchange port of the PBX

A.7 Sidetone, impedance matching and line loss

Poor (listener) sidetone causes problems if the ambient noise is high and the received speech level is low, i.e. the OLR value of the connection is at the limiting value. Therefore the ITU-T recommends that LSTR > 13 dB. To illustrate the consequences for the 2-wire impedance matching, we will study the case of a subscriber connected to a digital local exchange (or PBX) as shown in figure A.9.



NOTE: L = loss of the subscriber line. Lbr = Lst = weighted average of the return loss between the equ. sidetone balance impedance Zb and the input impedance Z of the line.

Figure A.9: Subscriber connected to a typical digital exchange

Let us put LSTR = 13 dB, the required minimum. With the designations used in figure A.9 we get

$$RLR = L + RLR(set) \qquad \text{(A.13)}$$

$$SLR = 7 + L + SLR(set) \qquad \text{(A.14)}$$

and by means of equ.(A.5) and (A.6) the lowest value for Lst:

$$(Lst)\min = 21 - D + 2L - (SLR + RLR) \quad \text{dB} \qquad \text{(A.15)}$$

For the typical case of (SLR+RLR) = 10, i.e. the telephone set is sensitivity regulated, and with D = 3 as is common for a good handset design, we thus get $(L_{st})_{min} = (8 + 2L)$ dB.

If one chooses lower values of SLR and RLR in order to accommodate possible connections with higher losses at the other end, the requirement on $(L_{st})_{min}$ rises accordingly.

Annex B: Noise aspects in understanding overall network performance

B.1 Introduction

This Annex is aimed to be a complement to subclause 6.7 with more detailed references to relevant ITU-T Recommendations and current investigations.

It is worth noticing that there is a new Question in ITU-T Study Group 12 "Noise aspects in the evolving network". The general task is to study what kinds of non-stationary noise can be identified in the actual network and what limits can be allowed for noise levels in wireless, satellite and multi-media communication. Moreover, as the mixed analogue/digital PSTN networks will interwork with ISDN networks, the noise levels and their statistical distributions will need investigation.

B.2 Noise uncorrelated with the signal

Noise uncorrelated with the signal may involve the following parameters of interest:

- circuit noise, weighted and unweighted (ITU-T Recommendations G.123 [13], G.153 [17], G.222 [20], G.223 [21], G.226 [22], G.228 [23], G.712 [24], Q.551 [29], Q.552 [30] and Q.553 [31]);

- noise introduced from power and traction lines (ITU-T Recommendations G.123 [13], G.151 [16], G.229 [120] and Supplement 13 [44]). Also new studies;

- "comfort noise", i.e. inserted artificially, noise contrast (GSM);

- impulsive noise, spurious transients (ITU-T Recommendations G.113 [35], P.11 [26] and P.55 [111]);

- noise bursts (For further study);

- single frequency noise (ITU-T Recommendations G.151 [16], G.712 [24], Q.552 [30] and Q.553 [31]);

- idle channel noise (ITU-T Recommendations G.712 [24], G.792 [123], Q.551 [29], Q.552 [30] and Q.553 [31]).

B.3 Signal correlated noise

Signal correlated noise may involve the following parameters of interest:

- quantizing distortion (ITU-T Recommendations G.113 [35], G.153 [17], G.222 [20], G.223 [21], G.712 [24], G.733 [122], G.792 [123], Q.551 [29], Q.552 [30] and Q.553 [31]);

- harmonic distortion and intermodulation (ITU-T Recommendation G.223 [21]);

- discrimination against out-of-band input signals (ITU-T Recommendations G.712 [24], Q.551 [29], Q.552 [30] and Q.553 [31]);

- spurious signals, spurious transients (ITU-T Recommendations G.113 [35], G.712 [24], Q.551 [29], Q.552 [30] and Q.553 [31]);

- sidetone distortion (ITU-T Recommendations P.11 [26] and Supplement No. 11 [41]).

B.4 Remarks regarding certain noise parameters

B.4.1 Impulsive noise

The effect of impulse noise depends on the rate of occurrence and the amplitude distribution of the individual impulses, see ITU-T Recommendations P.11 [26] and P.55 [111] for further information. However, this subject is under renewed study in ITU-T SG 12.

B.4.2 Bursts of errored bits, loss of cells in ATM, etc.

Voice Packeting Systems, Digital Circuit Multiplication Equipment, Digital Radio Systems and other systems using low bit-rate codecs are affected by these kinds of impairments. Further studies are needed. (However, a 6 ms break in a voice transmission, corresponding to the loss of a single cell in ATM, is hardly noticeable).

Annex C: Absolute and relative level

C.1 General

Transmission values for loss, gain and levels are expressed in decibels (dB) as a general principle. The basic unit "dB" is often extended with additional letters in order to distinguish between its use in different applications. The aim of this annex is to give a short description of the most common forms as used for transmission measurements at speech band frequencies as well as an introductory explanation of certain transmission planning applications.

C.2 The unit "dB"

This basic unit is mainly used for losses, gains, return losses etc., i.e. as a logarithmic ratio between two values which can be voltages, currents, powers, acoustic pressures etc. If the ratio is X for voltages, currents, pressures, the dB-expression is 20∗log(X). If the ratio is Y for powers, the dB-expression is 10∗log(Y).

C.3 The unit "dBm"

This unit with the additional "m" is used as a logarithmic measure of the "magnitude" P of an actual signal. The "dBm-value" of a signal is called its "absolute power level" or "absolute level".

The signal magnitude P used for signal characterization in speechband applications has the dimension of power, i.e. expressed in mW or mVA, and has by definition the form:

$$P = \frac{1\,000 \times (V)^2}{|Z(f_0)|} \qquad \text{mW or mVA} \qquad \text{(C.1)}$$

where:

V: the r.m.s. value in volts of the voltage across the test impedance Z which in the general case is complex and frequency dependent.

Z (f_0): the value of the test impedance in Ω at the (sinusoidal) reference frequency fo = 1 020 Hz.

The choice of this definition is based on three conventions:

In the first place it is practical to characterize the signal magnitude by a unit that has the dimension of power because this has been the practice for the special case of resistive terminations.

In the second place electronic circuits are designed to react on voltages, i.e. the open-circuit output voltage of, for instance, an amplifier depends only upon the voltage across its input terminals, irrespective of the input and output impedances of the amplifier. Thus, the "power" absorbed by the input impedance of the amplifier has no influence of how the signal is amplified. Hence the use of a constant impedance value in the denominator instead of a frequency-dependant impedance.

In the third place a sinusoidal signal with the reference frequency (1 020 Hz) the numeric value of P should be equal to the apparent power absorbed by Z, when this is complex, which is the same as the active power when Z is resistive.

Note that P is equal to the active power absorbed by the test impedance Z only when the latter is purely resistive and constant with frequency, for instance when Z = 600 Ω. Then P is measured in mW, otherwise in mVA. However, when Z is complex the value of P does not represent the apparent power absorbed by the test impedance at other frequencies than the reference frequency 1 020 Hz.

The definition for the so-called absolute power level L is:

$$L = 10 \log \frac{P}{P_0} \qquad \text{dBm} \qquad (C.2)$$

where:

P: the power in mW to be stated;

P_0: the reference value which is P_0 = 1 mW.

Sometimes the unit "dBm" is used in conjunction with a voltage level, referred to a voltage of 0,775 volts. The use of "dBm" in this application is only correct if the test impedance is 600 Ω resistive since 0,775 volts across 600 Ω results in the reference active power of 1 mW. This fact is important to remember if capacitive complex interfaces or test impedances are used.

C.4 The unit "dBr"

This unit is used to express the level relations for signals between points in a signal path, with the convention that one of the points is designated as a level reference point with the relative level 0 dBr.

More specifically, a sinusoidal reference signal of 1 020 Hz is thought to pass through the signal path under consideration with such an amplitude that its absolute level is 0 dBm at the 0 dBr point. The relative level in dBr at any other point in the signal path is then equal to the level (in dBm) that the reference signal has at that point (note that relative level designations must be used for both transmission directions).

What is the purpose of using relative levels?

It is immediately apparent that the differences of the relative levels between two points, which have the same level reference point, correspond to the loss or gain between those two points (at the reference frequency).

Moreover, relative levels are used to characterize the "power" handling capabilities of components (such as codecs) and equipment on the one hand and the expected levels of actual signals in the network on the other hand. This will be discussed in more detail in the following.

The "signal path under consideration", for which a specific 0 dBr reference point is designated, can encompass:

- a single component, such as an encoder or decoder;
- an equipment, such as a half-channel of a digital exchange;
- a circuit in the sense of the ITU-T definition, i.e. the fixed connection between two exchanges.

In the first two cases the "power handling capability" is the guiding principle for the allocation of a level reference point. For the third case the "expected absolute levels of actual signals" determines the choice of the level reference point.

The aim is of course to match the component performance to the requirement for the equipment performance which in turn should be matched to the actual range of signal levels. However, it is not always possible to achieve this exactly. For this and other reasons, the allocation of the 0 dBr reference point in the signal path may be chosen differently in the three cases above, i.e. when the component is considered alone, when it is considered as part of the equipment, and when the equipment is a part of the circuit. This means that the relative level designation for a certain point sometimes may differ in this three cases, a fact which should be remembered when discussing relative levels.

See also ITU-T Recommendations G.100 [114] and G.101 [115] for more discussions about relative levels.

Note that a so-called "level jump" may be introduced at the interconnection point between two (ITU-T) circuits. Thus, the loss or gain between two points belonging to two different circuits is not always equal to the difference in their (circuit) relative levels. Such an example is the case of the input and output relative levels of a digital exchange having no digital loss or gain pads. When the exchange is considered as an equipment, the difference between the (equipment) input and output relative levels gives the loss through the exchange because the two half-channels have the same level reference point. When the exchange is considered as a part of a connection, the two half-channels belong to two different (ITU-T) circuits which are interconnected "in the middle of" the switching matrix. The (circuit) input and output relative levels for the exchange, which are stated in the transmission plan for the connection, can differ from the specified (equipment) relative levels. This is because the (circuit) relative levels refer to two separate level reference points, each determined by estimation of expected signal levels in the two circuits (in general, however, the differences are not very large).

For the purpose of equipment parameter specification and transmission measurement, which is of interest here, the "power handling capability" is the governing factor for the choice of the 0 dBr level reference point. In this context, the digital 64 kbit/s PCM bit-stream is considered as having a relative level of 0 dBr, provided that there are no digital loss or gain pads in its path. Ideal encoders and decoders connected to the bit-stream are defined as having 0 dBr relative levels at their analogue ports when their clipping level for a sinusoidal signal lies at +3,14 dBm (A-law). The relative level for real encoders and decoders connected to the bit-stream is determined by means of the actual clipping levels in relation to the clipping levels of the ideal codecs.

When a digital loss or gain pad is included in the digital bit-stream, one has to make a choice of which side of the pad the bit-stream is to be assigned to 0 dBr. In the context of equipment specification and transmission measurement, it has been found most practical to apply a convention that a digital bit-stream never should be assigned a higher relative level than 0 dBr. This means that:

- A digital pad with L dB loss has the relative levels of 0 dBr at the input and -L dBr at the output;

- A digital pad with G dB gain has the relative levels of -G dBr at the input and 0 dBr at the output.

This is illustrated in figure C.1. Note that when a codec is to be tested alone, the digital bit stream is considered as being at 0 dBr at the codec output just as depicted in figure C.1a. If a configuration as shown in figure C.1b is to be tested, i.e. a codec in combination with X dB digital gain, the digital bit stream at the codec output becomes -X dBr. (Of course, the corresponding difference in relative levels also has to be observed at the analogue input of the codec).

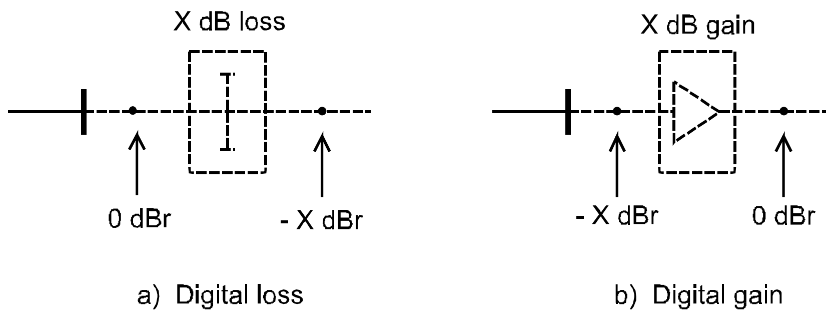
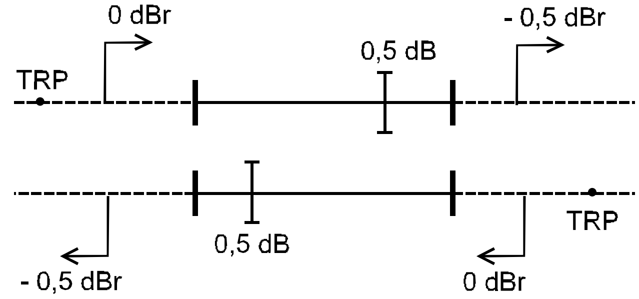


Figure C.1: Conventions, in the context of equipment specification and testing, for relative level of a digital bit stream when digital signal processing is applied

In the context of circuit relative level, a digital bit stream is usually, but not always, assigned 0 dBr. An exception to the general rule is shown in figure C.2. In this case, an analogue link is interposed between two digital links. (The analogue link needs 0,5 dB loss for stability reasons).



NOTE: TRP = the Transmission Reference Point, i.e. the circuit 0 dBr reference point.

Figure C.2: Example of (circuit) relative levels when an analogue link is interposed in a digital chain

Figure C.3 illustrates the fact that the loss between two points, belonging to separate circuits in a connection, is not necessarily equal to the difference in (circuit) relative level between those two points.

For stability reasons, there is a "level jump" of 1 dB at exchange 2. The loss of circuit 1 is 1 dB and for circuit 2 it is 0,5 dB. Thus, the loss of the connection is 1,5 dB, but the difference in (circuit) relative level between points 1 and 3 is only 0,5 dB. Likewise, the gain between points A and B will be 2 dB, i.e. a loss of -2 dB, but the difference in relative level is -3 dB!

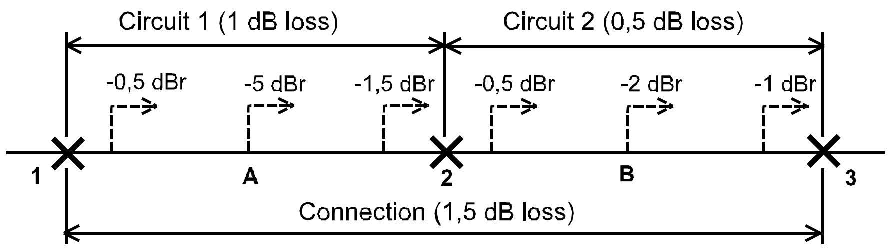


Figure C.3: Example of relative levels and losses in a connection consisting of two (ITU-T) circuits, where there is, for stability reasons, a "level jump" of 1 dB in exchange 2

C.5 The unit "dBm0"

When using an additional "m" and "0" (zero) with the basic "dB", the level under consideration is expressed as the absolute level (dBm) of the same signal that would be measured at the relevant 0 dBr level reference point.

This term is used in conjunction with transmission measurements to specify test levels and test results; the term also facilitates the comparison of the power levels of different signals by referring them to a common reference point, i.e. the 0 dBr reference point. Networks are often designed to carry different types of signals (speech, modem, fax, etc.) at different levels, expressed in dBm0.

C.6 The letter "p" in "dBmp" and "dBm0p"

The additional small letter "p" is derived from the French word "ponderé" for "weighted" and means that the considered value is a noise level, measured by a psophometer with a special noise weighting (called psophometric weighting) filter included, as described in ITU-T Recommendations O.41 [32] and P.53 [110].

C.7 The relationship between dBm, dBr and dBm0

The relationship between relative levels at interfaces and the resulting transmission loss or gain "L", is given by the formula

$L = L_i - L_o$

where L_i and L_o are the relative input and output levels at the interfaces.

The relation between the terms dBm, dBr and dBm0 can be expressed by the following formula:

dBm = dBm0 + dBr (general) (C.3)

dBmp = dBm0p + dBr (for weighted noise)

or

dBm0 = dBm - dBr (general)

dBm0p = dBmp - dBr (for weighted noise)

EXAMPLE 1: The test level for an interface with an input relative level of L_i = -2 dBr, should be -10 dBm0. To what absolute power level in dBm should the signal generator be adjusted?

dBm = dBm0 + dBr

= -10 + (-2) = -12 dBm.

EXAMPLE 2: The psophometrically weighted noise level at an interface with an output relative level of L_o = -7 dBr was measured to -69 dBmp. Does this value meet the requirement given with -65 dBm0p for this type of interface?

dBm0p = dBmp - dBr

= -69 - (-7) = -62 dBm0p.

The result shows, that the noise is outside the limit.

NOTE: Some modern test instruments provide as well an automatic adjustment of the correct absolute test level, as the necessary correction of received levels and displaying the results in "dBm0". In those cases the above given calculation can be avoided, however an additional adjustment (beside the test level itself) is required, to adapt the test instrument to the relative input and output levels of the test object.

C.8 Correction factors

Depending on the type of test instruments, auxiliary equipment and test objects, sometimes correction factors must be used, to either adjust the correct test signal level, or to obtain the correct test result. This mainly occurs in conjunction with capacitive complex impedances.

In practice, test instruments may be used with input/output impedances only 600 Ω resistive and consequently send levels or displayed results referred to 0,775 V. To provide the correct termination of test objects with complex impedances, auxiliary equipment called "impedance converter" are used. The principle of such an impedance converter is shown in figure C.4 in the application for sending and in figure C.5 for receiving.

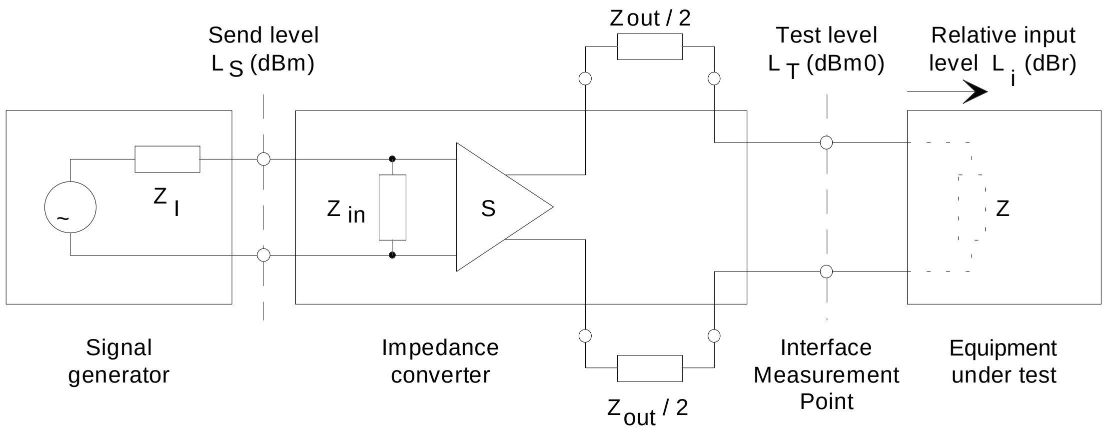


Figure C.4: Impedance converter in the sending path

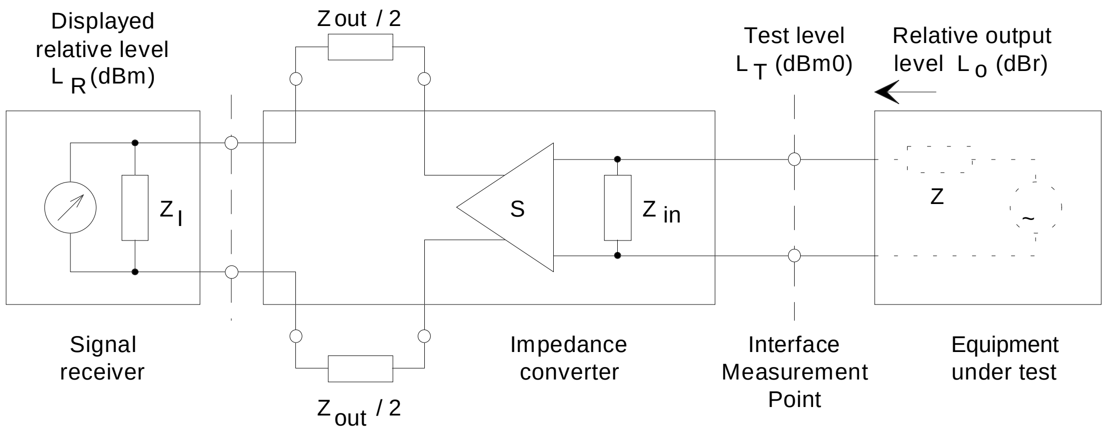


Figure C.5: Impedance converter in the receive path

The normal and advantageous design is, to obtain a power transfer ratio of 1 at the reference frequency 1 020 Hz, if terminated with the respective nominal impedances at input and output. In this case the gain "s" of the inserted amplifier is:

$$s = 6\,\text{dB} + 10\lg\frac{Z_{out}}{Z_{in}} \qquad \text{(C.4)}$$

This formula is valid for send and receive part of an impedance converter. It should be noted, that if Z_{out} or Z_{in} is a complex impedance, the modulus at the reference frequency 1 020 Hz must be used.

For impedance converters in the application with different complex impedances the gain "s" is normally only adjusted to 6 dB (power transfer ratio = 1 only, if $Z_{in} = Z_{out}$) and correction values are used.

a) Sending a test signal

In this application Z_{in} is exactly matched to the impedance Z_i of the signal generator (e.g. 600 W) and Z_{out} is the nominal interface impedance of the equipment under test (note, that the nominal value must be used, not the actual value).

To obtain the required test level L_T in dBm0 at the Interface Measurement Point, the necessary send level L_S in dBm of the signal generator can be calculated as follows:

$$L_S\,[\text{dBm}] = L_T\,[\text{dBm0}] + L_i\,[\text{dBr}] + 10\lg\frac{Z_{out}}{Z_{in}}$$

EXAMPLE 3: For an interface of an equipment with an input relative level L_i = -5 dBr and a nominal impedance Z = 842 W (modulus at 1 020 Hz for a 3-element complex impedance with 270W + 750W // 150nF) a test level of L_T = -10 dBm0 should be provided. What is the necessary send level L_S in dBm at a signal generator with 600 W impedance?

$$L_S\,[\mathrm{dBm}] = L_T\,[\mathrm{dBm0}] + L_i\,[\mathrm{dBr}] + 10\lg\frac{Z_{out}}{Z_{in}}$$

$$L_S = -10\,\mathrm{dBm} + (-5\,\mathrm{dBr}) + 10\lg\frac{842}{600}$$

L_S = -13,5dBm

b) Receiving a test signal

For receiving Z_{out} is exactly matched to the instrument impedance Z_i and Z_{in} provides the nominal termination of the equipment under test with the impedance Z.

To obtain the correct (received) test level L_T in dBm0 at the Interface Measurement Point, the displayed receive level L_R in dBm at the signal receiver must be corrected, using the following formula:

$$L_T\,[\mathrm{dBm0}] = L_R\,[\mathrm{dBm}] - L_o\,[\mathrm{dBr}] + 10\lg\frac{Z_{out}}{Z_{in}}$$

EXAMPLE 4: Assuming the same impedances for the test instrument (600 W) and the equipment under test (842 W) as in example 3, but with an output relative level of L_o = -7 dBr, what is the correct received test level L_T if the signal receiver readout is L_R = -50 dBm?

$$L_T\,[\mathrm{dBm0}] = L_R\,[\mathrm{dBm}] - L_o\,[\mathrm{dBr}] + 10\lg\frac{Z_{out}}{Z_{in}}$$

$$L_T = -50\,\mathrm{dBm} - (-7\,\mathrm{dBr}) + 10\lg\frac{600}{842}$$

L_T = -44,5dBm0

Annex D: Technical report on echo cancelling

NOTE: This is reprint of a draft T1A1.6 report ("Draft Technical Report on Echo Cancellers in Digital Networks", T1A1.6 Editorial Group [88], also given as a CCITT contribution, CCITT COM 15-33: "Technical Report on Echo Cancellers" [6]). It contains much useful information about echo elimination both in conjunction with voice as well as with data and facsimile transmission.

D.1 Scope, purpose and application

Echo cancellers are adaptive signal processors used to control echo; they are expected to replace echo suppressors in modern telecommunication networks. Echo cancellers are increasingly present on nearly every long distance connection and may be encountered singly or in tandem on a given link. The purpose of this technical report is to:

- explain the general principles of operation of echo cancellers;

- identify a limited set of application rules and the constraints under which echo cancellers operate;

- explain the relationship among the roles of the transmission planner of a Public Switched Telecommunications Network (PSTN), modem manufacturers, private network planners, and end users regarding the control of echo (from sources inside or outside the PSTN) and the associated terminal design considerations;

- identify how echo cancellers may affect the perceived quality of speech, the quality of voiceband data, as well as the performance of various signal processing and packetized circuit multiplication systems);

- identify both public and private network changes that may require additional study systems (such as of echo cancellers, to fully understand how these changes may impact the functionality of present echo cancellers;

- explain how new services, if accepted for implementation, could have an evolutionary impact on echo canceller design.

D.2 Historical background

D.2.1 PSTN transmission planning

In the telephone network, the access line is typically a 2-wire facility between the customer premises and the switch, while the transmission facilities between the switches are typically 4-wire on long connections. At the 4-wire-to-2-wire conversion point, which typically occurs in a switch line card, a perfect impedance match cannot be achieved and thus a return signal, referred to as echo, results. Therefore, one of the major concerns of the PSTN transmission planner is to ensure adequate echo control to provide satisfactory transmission performance.

For low-delay connections, echo is controlled by the insertion of appropriate transmission path losses, as defined in ANSI T1.508 [153]. Longer delay connections need echo control devices. It is the PSTN transmission planner's role to design PSTNs so that the echo control devices installed provide adequate echo control at the 4-wire-to-2-wire conversions that occur in the PSTN, and to ensure that the customer obtains satisfactory transmission performance.

In the past, echo suppressors were used to control echo in long distance networks. Today, however, the echo canceller is the device of choice. The following two subsections summarise the reasons for the use of echo cancellers instead of echo suppressors in modern telephone networks. (See references [73], [77], [93], [99] and [100]).

D.2.2 Echo suppressors

The principle of echo suppressors is well-known; it is summarised as follows: When speech is detected on the receive path, a very high attenuation is inserted in the send path. When double-talk is detected, the send path is closed and a receive loss is inserted in the receive path. Thus, during double-talk, there is no echo suppression, but the echo is much more attenuated than the direct speech. Other refinements are possible, as indicated in CCITT Recommendation G.164 [18].

Many problems can occur in the operation of echo suppressors; this is because the decision as to which end is talking and which is listening is based essentially on the transmission levels. If the level of the echo is high and the level of direct speech is low, speech could be mutilated and/or it could be difficult to distinguish between single-talk and double-talk. This could also be the case at the beginning or at the end of a speech burst.

The problems are compounded on long-delay transmission paths, because the pattern of conversation is usually changed. In addition, the cascading of echo suppressors is not recommended. In the case of voiceband data, a 2 100 Hz tone is specified to permit disabling of the echo suppressor before the beginning of data transmission; this is for two reasons:

- to avoid insertion losses for modems with a secondary channel;

- to avoid delays due to hangover at turnarounds, thereby increasing the throughput.

Facsimile is a special case. Even if an echo suppressor is disabled by the 2 100 Hz tone, it may be re-enabled during a facsimile transmission. The tone disabler hangover time of an echo suppresser is specified as 250 +/- 150 ms in para 5.7 of CCITT Recommendation G.164 [18]. Therefore, periods of silence greater than 100 ms at the echo suppressor may cause the echo suppressor to be re-enabled, while periods greater than 400 ms cause it to be re-enabled. During a facsimile call, there are a number of silent periods that may be long enough to permit the re-enabling of the echo suppressor. In addition, some facsimile manufacturers have chosen to exceed the signal separation intervals specified in CCITT Recommendation T.30 [139]; therefore, echo suppressors may be re-enabled.

Enabled echo suppressors may distort the facsimile signals. One type of distortion is the truncation of fast turnaround signals. Typically, the echo suppressor operates in a single-talk mode, so that when a signal arrives at the receive port, the suppression switch is activated and remains in that state until no signal arrives for a certain time. The recommended hangover time associated with each state transition is in the range of 24 ms to 36 ms (see NOTE), as specified in table 4 of ITU-T Recommendation G.164 [18]. The suppression hangover time guards against echo stored in the local echo path.

NOTE: Analogue echo suppressors conforming to CCITT Recommendation G.161 [118] are still in service; they have a suppression hangover time of 40 ms to 75 ms (see table 4 of ITU-T Recommendation G.161 [118]). Accordingly, a signal may be mutilated if it reaches the transmit port before 40 ms to 75 ms.

Note that the CClTT has declared CCITT Recommendation G.161 [118] to be obsolete after 1992.

Now, ITU-T Recommendation T.30 [139] specifies that the guard time between ITU-T Recommendation V.21 [140] and ITU-T Recommendation V.29 [145] transmission should be 75 ± 25 ms. If a return signal from the local facsimile machine (within a V.21 message-response sequence or a V.21/V.29 sequence such as a confirmation to receive (CFR) followed by training) reaches the echo suppressor transmit port within 24 ms to 36 ms of the termination of the signal at the receive port, the persistence of echo suppression insertion losses or open-circuit condition may introduce an attenuation. As a result, the echo suppressor mutilates the initial portion of that fast turnaround signal. When this signal is part of the training/training check signal, training might be disrupted and rate fallback ensues, or in a worse case, the call is terminated.

Similarly, an enabled echo suppressor may block a low-level secondary channel signal. If the level of that signal is high enough, the suppressor may enter the double-talk mode, in which a receive loss is inserted. The result is a reduction in the levels of both the transmit and the receive signals, if echo suppressors are at both ends of the connection and are both in the double-talk mode.

Finally, for certain combinations of propagation times and insertion losses, listener echo may cause the 2 100 Hz tone to persist long enough to disable the echo suppressors. This echo may then contribute to the degradation of the image quality by reducing the signal-to-noise ratio during page transmission.

Prior to ITU-T Recommendation V.32 [146], most 2-wire modems used frequency division to provide duplex operation (i.e, different carrier frequencies were used for each direction of transmission). In the early 1980s, data showed that some echo cancellers did improve the operation (i.e., reduce or eliminate bit errors) for low-speed modems designed according to CCITT Recommendations V.21 [140], V.23 [141], V.26 [142] (alternative B), V.27 [143], and V.29 [145] (see also reference [76]). Therefore, it was accepted that these modems benefited from an active echo canceller and a disabled echo suppressor. Accordingly, in 1984, the CCITT modified CCITT Recommendation G.165 [19] to recommend that echo cancellers be disabled with a 2 100 Hz tone with phase reversals.

Recently, preliminary data have indicated that certain combinations of modems/echo cancellers, in various simulated network configurations, exhibit degraded performance when the echo cancellers are enabled (see reference [72]).

V.32 modems, in contrast, use the same band of frequencies in both directions and achieve duplex operation through the use of an integrated echo canceller. The echo canceller integrated in this voiceband data modem should not be confused with the network echo cancellers that conform to CCITT Recommendation G.165 [19], because the performance requirements for each are very different.

D.2.3 Echo cancellers

Echo cancellers are voice-operated devices that use adaptive signal processing to reduce or eliminate echoes. Echo cancellers are placed in the 4-wire portion of a circuit, and reduce (or cancel) the echo by subtracting an estimate of the echo from the returned echo signal. Echo cancellers may operate on a single circuit or on a multiplexed facility, e.g., echo cancellers operate on a 64 kbit/s speech facility that is multiplexed into a primary rate link. Echo cancellers are designed to:

- cancel linear echo path signals;
- refrain from cancelling the echo when requested to do so by an in-band disabling signal;
- return to an operational mode after being disabled when the in-band signal power level drops below a specified level for a specified period of time. This design allows some networks to transport voiceband data on the same speech channels. It also allows the echo canceller to re-enable during a voice call after it has been turned off erroneously (talkoff).

Echo cancellers are characterised by whether the interface path is analogue or digital, and/or whether the subtraction of the echo is by analogue or digital means. This technical report is limited to echo cancellers that have a digital input and digital subtractors. CCITT refers to this type of echo canceller as a Type C echo canceller.

Echo cancellers have the following main advantages over echo suppressors:

- send path transparency is improved;
- hangover introduces fewer impairments;
- there is no receive insertion loss;
- echo cancellation continues during double-talk;
- cascading is possible (for well designed echo cancellers).

Some echo cancellers are optioned to disable on the 2 100 Hz tone specified in CCITT Recommendation G.164 [18] for echo suppressors, and some are disabled with a 2 100 Hz tone with periodic phase reversals of 180°+/- 25°, as specified in CCITT Recommendation G.165 [19] for echo cancellers. Use of the G.165 tone is intended to allow echo cancellers to be disabled independently of echo suppressors.

Most modem manufacturers feel that network echo cancellers should be disabled for modems with integrated echo cancellers (e.g., ITU-T Recommendations V.32 [146] and V.34 [51]), because an active network echo canceller operating in conjunction with the integral echo canceller in the modem may cause undesirable phenomena under specific but unlikely circumstances. Some of these cases are:

- the echo canceller incorrectly identifies the near-end signal as an echo and attempts to cancel it;

- when there is frequency offset in the tail circuit, the echo canceller injects bursts of reinforced echo interspersed with quiet periods;

- although neither case is likely, it was decided that the onus for making the decision to disable the network canceller should rest with the end users. Modem manufacturers had to rely on a unique technique to disable echo suppressors and echo cancellers.

Historically, manufacturers of modems with integrated echo cancellers have designed their modems to disable network-based echo cancellers. These modems disable network-based echo cancellers using the disabling tone specified in CCITT Recommendation G.165 [19]. Modem-based echo cancellers must accommodate three types of echoes simultaneously:

1) near-end echo;

2) far-end echo, and

3) any echo generated between the near-end and the far-end.

Because the range of tail length needed for each case varies widely, three echo cancellers may be needed.

D.2.4 Responsibilities of modem manufacturers and end users

It is the responsibility of the modem manufacturers and end users to understand the characteristics of the network-based echo canceller fully and decide whether the echo cancellers should be enabled or disabled. If the modem manufacturers and end users decide that the network-based echo canceller functionality should be disabled, they must ensure that the terminal uses the appropriate approved methods, defined in CCITT Recommendations, to disable cancellers. Additionally, it is the end user's responsibility to ensure that terminals and private networks are designed to operate in a fashion compatible with the PSTN network-based echo cancellers. For example:

- digital telephone sets are expected to control their own echoes (the PSTN network is not responsible for cancelling acoustic echoes);

- terminals and private networks should be designed to provide circuit extensions compatible with the design intent of the PSTN, e.g., echo paths outside the network should be linear or the terminal should control its own echo;

- either the delay of the terminal or private network should be within the operational limits of the network-based echo canceller, or the terminal/private network should control its own echo.

D.3 Reference connection

D.3.1 Echo control reference model

Figure D.1 shows the reference model for echo cancellers; it is adapted from CCITT Recommendation G.165 [19].

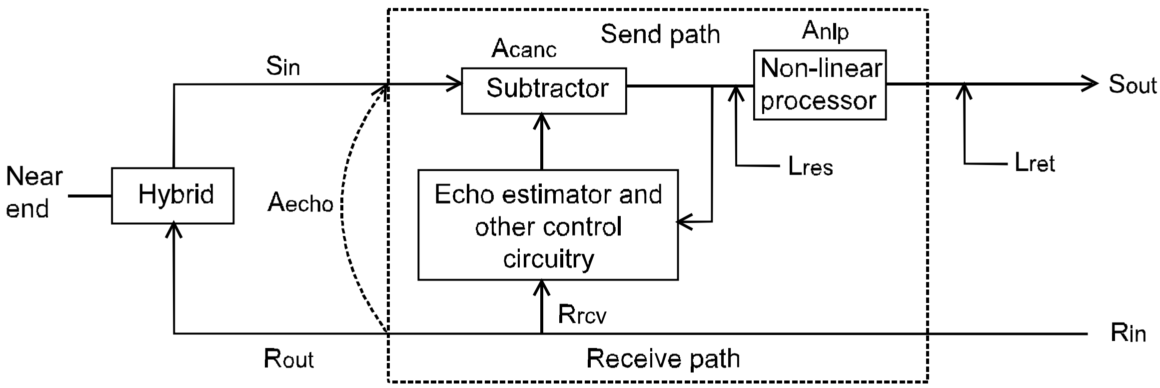


Figure D.1: Echo canceller reference model

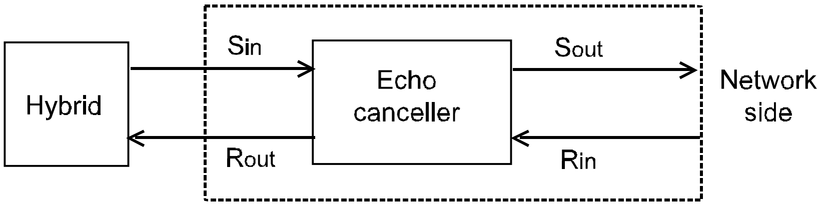


Figure D.2: Designations of the echo cancellers ports

D.3.2 Terminology

The following terms are defined in a manner consistent with both the 1992 revision of CCITT Recommendation G.165 [19] and the model in figure D.1. The terms, listed in alphabetical order, should be read while referring to figure D.1.

Cancellation (A_{canc}): The attenuation of the echo signal as it passes through the send path of an echo canceller. This definition excludes any nonlinear processing on the output of the canceller to provide further attenuation.

Convergence time: For a defined echo path, the interval between the instant a defined test signal is applied to the receive-in port (R_{in}) of an echo canceller with the estimated echo path impulse response initially set to zero, and the instant the returned echo level L_{ret} at the send-out port S_{out} reaches a defined level.

Dispersion time: The time required to accommodate band-limiting, multiple reflections and transit time in the hybrid.

Echo path delay (end delay or tail length): The duration of the echo to be cancelled. In a single echo path, the echo path delay td is the sum of pure delay and dispersion time. If there are multiple echo paths, the overall echo path delay is the maximum of the individual echo path delays.

Echo loss (A_{echo}): The attenuation of a signal from the receive-out port (R_{out}) to the send-in port (S_{in}) of an echo canceller, due to transmission and hybrid loss, i.e., the loss in the echo path.

Near-End Speech Threshold (NEST): The minimum attenuation of the signal between port R_{out} and port S_{in} for the echo canceller to declare that only echo is present.

Nonlinear processing loss (A_{nlp}): The attenuation of the residual echo level from the output of the canceller that is due to a nonlinear processor (NLP) placed in the send path of an echo canceller.

Pure delay (tr): The time taken by a signal sent out on port R_{out} to reach the port S_{in} inherent in echo path of the transmission facilities. In this case, the transit time directly across the hybrid is assumed to be zero.

Residual echo level (L_{res}): The level of the echo signal that remains at the send-out port of an operating echo canceller after imperfect cancellation of the circuit echo. It is related to the receive-in signal LR_{in} by

$$L_{res} = LR_{in} - A_{echo} - A_{canc}$$

Any nonlinear processing is not included.

Returned echo level (L_{ret}): The level of the signal at the send-out port of an operating echo canceller that will be returned to the talker. The attenuation of a nonlinear processor (NLP) is included, if one is normally present. L_{ret} is related to LR_{in} by

$$L_{ret} = L_{res} - A_{nlp} = LR_{in} - (A_{echo} + A_{canc} + A_{nlp})$$

If nonlinear processing is not present, note that $L_{res} = L_{ret}$.

D.4 Application rules and operational constrains

D.4.1 Public network transmission planning

D.4.1.1 Transmission levels

The network transmission planner is expected to implement a valid loss plan in the evolving digital PSTN to ensure that appropriate transmission levels exist at the various A/D conversion points (see ANSI T1.508 [153]). By having such a formal plan, pulse code modulation (PCM) overload distortion is avoided and signal levels allow the echo canceller to operate as per its design intent.

Guidance for transmission levels can be found in the CCITT G.100-series of Recommendations for PSTNs that utilise analogue accesses and for connections from digital cellular networks. Mu-law encoders must be consistent with CCITT Recommendation G.711 [62]. For PSTNs with digital access, guidance for terminal design can be found in CCITT Recommendation P.31 [39]. Note that while these Recommendations provide guidance as to appropriate design levels, FCC Part 68 and Canadian Standard CS-03 provide requirements on maximum signal levels entering the network.

D.4.2 Delay considerations

As previously mentioned, conversion from the 4-wire toll network transmission facilities to 2-wire loop plant facilities must be made on all long connections. On these connections, it is the impedance mismatch at the hybrid that causes reflections of the incident signal at the 4-wire interface to occur (see figure D.1). The amount of attenuation of the reflected signal S_{in} relative to the incident signal, R_{out}, is known as echo return loss (A_{echo}). Because loops vary in composition, e.g., their length varies and they may be loaded or unloaded, a perfect balance cannot be obtained. Based on empirical data, it is commonly accepted that the average A_{echo} should be considered to be approximately 11 dB. For those loops in which a poor impedance match is obtained, the reflections (talker echo) can become noticeable and objectionable when the delay between two telephones is greater than about 16 ms (32 ms round-trip). (See CCITT Recommendations G.131 [14] and G.114 [36] for guidance in this regard). It is the transmission planner's responsibility to determine at what point, i.e., for what delay threshold, a network echo control device will be implemented. This is a business decision that requires a balance between performance and cost.

NOTE: If an appropriate transmission plan is not implemented, echo may still occur in a circuit equipped with echo cancellers.

D.4.2.1 Echo return loss

The Near-End Speech Threshold (NEST), or Double-Talk Detection Threshold (DTDT), is the level at hich the echo canceller declares the presence of near-end speech, i.e., the occurrence of double-talk, and stops its adaptation process. In other words, double-talk is declared when:

$$R_{out} - S_{in} < NEST/DTDT$$

For example, when the NEST/DTDT of an echo canceller is provisioned for 6 dB, the echo canceller declares near-end speech and stops its adaptation process if $R_{out} - S_{in} < 6$ dB.

It is important that the NEST/DTDT value be provisioned such that the A_{echo} > NEST/DTDT. For example, when the echo canceller is provisioned for NEST/DTDT = 6 dB, the echo canceller works properly with a 4-wire circuit path whose $A_{echo} > 7$ dB. However, if the hybrid has $A_{echo} < 6$ dB, the echo canceller assumes that the echo at the S_{in} is a near-end speech, and stops its adaptation process. The end result is the presence of echo on the S_{out} path; the echo canceller does not cancel the echo.

When the echo return loss (A_{echo}) is less than a provisional threshold, the echo return loss of the circuit must be increased through level adjustments. It is the transmission planner's responsibility to ensure A_{echo} is greater than the NEST/DTDT for which the circuit is provisioned.

D.4.3 Provisioning or the tail length

The link from the canceller to the hybrid is often referred to as the "tail" of the circuit The delay of the echo to be cancelled is determined by specifying the "tail length" of the canceller. To specify this tail length correctly, it must be remembered that some of the received power at port R_{out} is reflected by the hybrid resulting in echo at port S_{in}. The time it takes the signal at R_{out} to travel from the echo canceller to the hybrid and return to the echo canceller at port S_{in} must not exceed the provisioned tail length; otherwise the echo cancellation process will not work properly. This time must include round-trip propagation time delay over the transmission media, all intermediate equipment, and the dispersion due to the transmission characteristics of the circuit. This dispersion increases the effective duration of the impulse response of the echo canceller.

It is the transmission planner's responsibility to ensure that echo cancellers are implemented in such a way that their tail delay is not be exceeded, on normal network connections, so that echo cancellation occurs. Cooperation between the interexchange carriers and the exchange carriers is required to ensure that the planner can properly implement cancellers. In an addendum to T1.508 [153], the T1A1 committee addresses the allocation of a maximum expected value of delay for exchange carriers.

D.4.4 End user/manufacturer/private network transmission planning

For convenience, the term "end user/manufacturer/private network planner" is used synonymously with "private network transmission planner".

D.4.4.1 Transmission levels

The private network transmission planner is expected to implement equipment that is consistent with the network transmission loss plan. Guidance is available in the form of CCITT Recommendations. Further, the private network transmission planner is expected to meet the requirements of FCC Part 68 and Canadian Standard CS-03.

D.4.4.2 Delay considerations

The private network transmission planner, like the network transmission planner, should make a conscious decision about how to control talker echo, and at what delay threshold to implement an echo control device in the private network. Note that if the private network connects to the PSTN on a 4-wire basis, the echo generated by the 4-wire to-2-wire conversion may be cancelled by the network-based echo canceller. However, if the private network connects to the PSTN on a 2-wire basis and then converts to 4-wire for carriage, the private network transmission planner must consider how to handle the echoes generated at the 4-wire-to-2-wire conversion points in the private network.

D.4.4.3 Echo return loss

It is the private network transmission planner's responsibility to ensure that A_{echo} is greater than the NEST/DTDT for which the circuit is provisioned (see subclause D.4.2.1).

D.5 Effect of cancellers on voice and data services

Network-based echo cancellers are present on connections that experience long delays. They must be designed to allow a speech channel to support voiceband data, including facsimile. This means that they must retain the capability of being disabled upon an appropriate request from customer terminal equipment. However, the modem manufacturer is responsible for determining if network-based echo cancellers should be enabled or disabled.

D.5.1 Interaction with voiceband data

Full-duplex data transmission in the voiceband can occur, depending on the modem modulation scheme. New modulation schemes are being introduced, and manufacturers must determine the optimal state in which the echo canceller should be when the modem is operating, i.e., if the canceller should be enabled or disabled, or whether the call should be routed on a connection that never has an echo canceller functionality present.

D.5.2 Interaction of echo control with facsimile transmission

The designers of facsimile terminals generated these terminals with the understanding that network providers were installing network-based echo control devices as per CCITT Recommendations G.161 [118], G.164 [18], and G.165 [19]. Thus, PSTN network planners were expected to continue to evolve the network in such a way that it would not knowingly prevent the continued carriage of a permissive voiceband data/facsimile service.

Although facsimile machines may transmit a G.164 disabling tone at the beginning of a call, there is no requirement to guarantee that the power of in-band signals will continue to hold the echo control devices in the disabled state for the duration of the call. Echo control devices conforming to CCITT Recommendations G.161 (analogue echo suppressors) [118], G.164 (digital echo suppressors) [18], and G.165 (echo cancellers) [19] are designed to re-enable when the signal level drops below a predefined threshold for a predefined period of time, once the call is in progress. The reason for this is that echo control devices conforming to CCITT Recommendations G.161 [118], G.164 [18], and G.165 [19] are designed to become re-enabled if no signal energy is present in both directions of signal transmission for a period greater than 100 ms (minimum) to 400 ms (maximum) (see para 5.2 and 5.5 of CCITT Recommendation G.164 [18]).

The V.27ter [144] modulation scheme employed by CClTT Recommendation T.30 [139] is protected against the mutilation of the training sequence by echo suppressors (by using an unmodulated carrier prior to the training signal). In contrast, the V.29 modulation scheme is not protected. Some implementations are based on proprietary solutions to this problem (most notably the addition of an unmodulated carrier prior to V.29 transmissions of the same format as that used during V.27ter transmissions). Unfortunately, these schemes are not universally recognisable by terminals produced by different modern manufacturers. Thus, if the guard time between V.21 and V.29 transmission from the facsimile machine exceeds the T.30 time limit of 75 +/- 25 ms, it is possible that an echo suppressor will be re-enabled. In this case, the initial portion of the training check sequence could be mutilated, preventing the connection establishment.

The presence of echo can interfere with facsimile transmission in two ways:

- the echo could be misinterpreted as a T.30 protocol message and then interrupt the handshake between the two ends machines. This is particularly important if the facsimile machines are not protected against echo;

- the echo can reduce the S/N ratio necessary for the good transmission of image data.

Echo could be present for the following reasons:

- echo suppressors are disabled (to avoid errors in voiceband transmission). As explained earlier, enabled echo suppressors may cause errors in voiceband data transmission. However, it may be preferable to keep them enabled during facsimile transmission;

- if echo cancellers are disabled according to the procedures of CCITT Recommendation G.164 (2 100 Hz tone) [18], then, depending on the propagation delay and the response time of the facsimile machines, echo could be present during the initial handshake. This could disrupt the establishment of the call. This imposes a limit of 400 ms on the time during which no energy could flow in either direction for the echo control device to re-enable. If these echo cancellers remain disabled, the echo of the V.21 signal may confuse the facsimile machine at the other end and/or confuse the facsimile demodulator of the network packetized circuit multiplication equipment (PCME)/ digital circuit multiplication equipment (DCME). The image quality may be affected as well;

- echo cancellers that respond to the G.165 disabling tone are not be affected by the 2 100 Hz tone without phase reversal, and remain enabled.

Other vulnerable instances during the connection are when handshakes are exchanged between pages. Disabled echo cancellers could allow echo at these instances; enabled echo cancellers, in contrast, control echo, including listener's echo.

Under some conditions, echo cancellers disabled using the G.164 procedures (2 100 Hz) may affect the connection establishment or the quality of facsimile transmission because they may be disabled inadvertently by the called station identification (CED) tone; hence, echo control does not function as expected. Echo cancellers disabled using the G.165 procedures (2 100 Hz with periodic phase reversals) do not suffer from the problems described previously, because they are reliably disabled, and therefore do control the echo in a facsimile connection.

It should be noted that a number of echo cancellers already deployed in the PSTN are not able to eliminate completely short echo bursts that could occur while the canceller is reconverging after transitions between the narrow-band signals, such as the CED tone or the V.21 high-level data link control (HDLC) handshake, the wideband image signals (e.g, V.29 or V.27ter signals), and again, narrow-band signals. In the future, it still will not be possible to guarantee that all echo cancellers will be able to avoid this problem.

NOTE: This paper does not discuss explicitly the case in which there is one echo canceller on one side of the connection and another on the other side; this "mixed case" can be deduced from subclauses D.2.2 and D.2.3.

Current CCITT Recommendations imply that echo cancellers should be enabled during facsimile transmission. Generally, echo suppressors do not provide the same level of performance for speech, voiceband data, or facsimile. Enabled echo suppressors could cause failures due to clipping and/or mutilation of the training check sequence, thereby preventing the establishment of the facsimile connection. However, it may be better to enable echo suppressors during facsimile transmission to protect against both talker and listener echoes and avoid their interference with facsimile at connection establishment and/or during image transmission.

The main conclusion is that it is better to use echo cancellers that are disabled according to the G.165 procedures.

D.6 High-level speech

D.6.1 Introduction

A number of sources could produce high speech levels in the network. In hands-free telephones, for example, the microphone may allow high speech levels to be generated. With this perspective in mind, CCITT Recommendation G.165 [19] was modified in 1992 to include an overload test (Test No. 8) at levels exceeding 0 dBm0 (the provisional values for the test are +3,05 dBm0 and +3,25 dBm0). The presence of high speech levels may cause increased nonlinearities that would degrade the performance of some echo cancellers, especially echo cancellers that have not been implemented in a fully digital manner. Another area in which high signal levels may cause difficulty is in the double-talk detection and nonlinear processor control circuits. These are discussed in the following two subclauses.

D.6.2 Double-talk detection

The performance of echo cancellers is very dependent on the double-talk detection algorithm used. For example, if double-talk is not recognised quickly, the near-end speech masks the residual echo that is used to update the impulse response model of the echo canceller. The effect of double-talk detection in the presence of high signal levels should be considered for study; possible new echo canceller requirements for echo canceller design may result.

D.6.3 Effects of a non-linear echo path

The theory of echo cancellation assumes that the echo path is linear. Therefore, it is critical that no clipping or nonlinear distortion occur in the echo path between R_{out} and S_{in}. If any clipping does occur, it is important that it be slight, infrequent, and that it occurs only during double-talk conditions. Otherwise, the environment should be corrected, e.g., frequency offset should be removed or implementation of an acceptable transmission plan should be ensured.

One potential source of problems with high-level speech stems from the resultant nonlinearities in the echo path. For optimal echo canceller performance, it is essential that the signal fed into the echo canceller's R_{in} port be linearly related to the signal at the echo canceller's S_{in} port. If any nonlinear distortion of high-level speech occurs, the distortion must occur before it is used by the echo canceller so that the same clipped signal is sent to the R_{out} port. However, echo-canceller performance may still degrade if the tail path is not linear.

Some echo cancellers use the signal at R_{in} as its internal received signal R_{rcv}, and also pass R_{in} to the R_{out} port. This is acceptable provided that there is no clipping or other nonlinear distortion of one signal leg that does not occur with the other. Otherwise, the echo path does not appear to be linear to the echo canceller and, consequently, performance suffers.

Additionally, clipping or other nonlinear distortion should not be "added" to the signal at the S_{in} port. This is most important when: (1) echo is present only at the S_{in} port, or (2) both echo and near-end speech are present and the double-talk detector has not been triggered, since clipping (distorting) one affects the other.

D.6.4 Guidelines for Rout usage in echo-cancellers

The configuration in which the same signal feeds both R_{in} and the echo path may result in degraded performance if R_{out} is not digitally equivalent (bit for bit) to R_{in} under all signal conditions. The signal R_{rcv} used internally by the echo canceller after passing through the R_{in} port can be used as the source signal for the echo path. Therefore, it is recommended that R_{out} (which is used to drive the echo path) should be digitally equivalent to R_{rcv}.

D.7 Network and service evolutionary considerations

D.7.1 Communication with switches

Echo cancellers are signal processing devices that historically have been applied on a dedicated trunk basis within the telephone network. As migration from channel-associated signalling to common channel signalling progresses, trunk efficiency could be improved if the echo cancellers could be enabled or disabled using class of bearer service information. To implement such an arrangement using non-switch-associated cancellers, a communication capability is required between the switches and the echo cancellers.

The need for the standardisation of such a signalling link requires further study.

D.7.2 Bit transparency of echo cancellers

CCITT SG XV has amended CCITT Recommendation G.165 [19] to make it clear that a 2 100 Hz disabling tone with phase reversals tone shall cause the echo canceller to disable and provide an analogue clear-channel signal path. In other words, a tone between 300 Hz and 3 400 Hz shall pass with its power level and frequency unaltered through the echo canceller, but 64 kbit/s transparency is not guaranteed. (See para 3.3 of CCITT Recommendation G.165 [19], 1992 revision). It should be noted that 64 kbit/s transparency is achievable and is implemented in some echo cancellers, but to remain in that state, the in-band power level must remain above a predefined power level.

If cancellers are to be applied to trunks and disabled by use of a "switch to echo canceller" signalling channel, the canceller should support a 64 kbit/s clear channel capability, if such capability is to be provided.

D.7.3 Non-linearities in tail circuit

Two issues are related to the introduction of nonlinear signal processing techniques in the PSTN: 1) the occurrence of voice compression in the echo path, and 2) the occurrence of digital insertion losses.

- with the increasing use of voice compression in the public and private voice networks, specifically 32 kbit/s adaptive differential pulse code modulation (ADPCM) (ANSI T1.303 [154]), the occurrence of a voice compression codec in the echo path becomes more likely. The effect of voice compression techniques as a nonlinearity affecting canceller performance is for further study;

- with the increased use of digital techniques in processing voiceband signals, digital insertion losses are being implemented increasingly in digital pads. However, improperly designed digital pads may add substantial nonlinearities to the transmitted signal, including the returned echo signal, therefore degrading the canceller performance. Such digital padding typically occurs in PSTN end offices when they act as a host to a digital remote line module as well as in customer premises equipment (CPE), such as private branch exchanges (PBXs). The need to maintain linearity in digitally padded signals should be recognised.

D.7.4 Voice compression between tandem cancellers

The use of voice compression as part of the voice transmission path could also affect circuits that use tandem cancellers (see figure D.3). Consider the circuit in which tandem cancellers are in place, and voice compression is used only between the two cancellers. Although the canceller closer to the hybrid would not be affected, the canceller on the network side would see a tail circuit with a nonlinearity, as described in subclause D.6.2. The performance of the tandem still may be acceptable if the canceller closer to the network remains stable and maintains a return loss enhancement. Theoretically, the canceller on the network side never sees an echo because the canceller on the distant end has removed it. Therefore, the cancellers on the network side should be removed effectively from the connection.

The conditions under which the performance is not degraded is a candidate for further study.

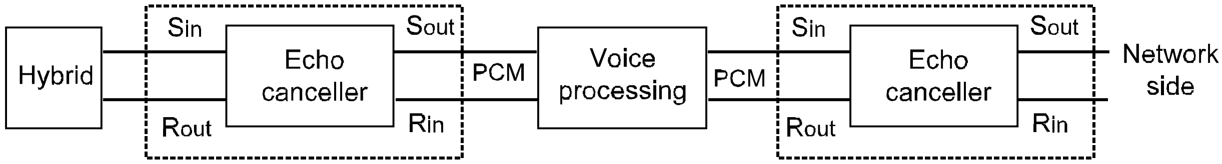


Figure D.3: Voice processing between two echo cancellers

D.7.5 Tandeming of echo cancellers

It is generally accepted that properly designed echo cancellers can be tandemed with little or no penalty in performance. In CCITT Recommendation G.131 [14], Rule B indicates that G.165 echo cancellers can be connected in tandem without echo performance degradation (see subclause 2.3.2.1.1 of CCITT Recommendation G.131 [14]). With the increasing use of dynamic routing and special features such as call forwarding, and because of the long delay introduced by low bit-rate speech coders in cellular applications, it is very likely that some connections have more than one echo canceller.

Subjective tests of some echo cancellers verify that tandeming poses no problerns under most conditions. However, reports have suggested that other echo cancellers cannot be tandemed without problems. In these cases, it is imperative that the PSTN and/or private network transmission planner ensures that echo cancellers that cause undue performance degradation when tandemed are not allowed to operate in a tandem mode.

The data suggest that improper design of some of the auxiliary circuits, such as nonlinear processors, could cause problems when the echo path delay for one of the echo cancellers in tandem exceeds its tail capacity. For example, in some echo cancellers, the NLP may operate at inappropriate times during double-talk. This occurs when the hangover time in the NLP circuit does not match the echo path delay characteristics.

To illustrate, assume that the NLP algorithm is designed to operate on the basis of the NEST/DTDT value. In the case where the tail delay capacity of an echo canceller is exceeded, the echo arrives later than the "expected" time. As a result, the comparison is in effect between power levels of a later far-end speech burst and an unrelated near-end speech burst. Based on this scenario, clipping can occur. However, it is reasons like these that make it important that PSTN and private network transmission planners ensure that the delays of echo canceller tails are never exceeded, unless additional echo control measures are taken inside the private network.

This problem is mitigated since it only occurs during double-talk, and most situations involving tandeming of echo cancellers do not include many cases in which the tail path delay capacity is greatly exceeded. Finally, with some adjustments to the time constants of the NLP, partial improvements can be made.

It has been observed that if an echo canceller converges too quickly, it can have annoying side effects if it is used in a situation where its tail length capacity is exceeded (such as sometimes occurs with tandem echo canceller operation). Therefore, the tail delay of an echo canceller should be 4 ms to 6 ms larger than the rnaximum expected network delay, as estimated from table 1 of CCITT Recommendation G.114 [14]. This takes into account the effect of dispersion. For example, to take into account a maximum pure delay of 44 ms, a 48 ms canceller could be selected.

D.7.6 Convergence speed

Some echo cancellers generate noise in trying to continuously adapt to the echo path. This may be related to adaptation speed. The effect is very noticeable and annoying, especially during double-talk, when the adaptation process is suspended. Generally speaking, as the speed of adaptation is increased beyond the optimum speed, the accuracy of the transfer function after adaptation becomes poorer. High speed of convergence is desirable for initial acquisition, while lower convergence may be needed for subsequent tracking, since the echo transfer function changes very slowly.

D.7.7 Acoustic echo control

Acoustic echo control is becoming an important issue due to the increase in hands-free telephone sets. Although there is some commonality between issues encountered for acoustic echo cancellation and network echo cancellation, there are also many differences. The issues of level points, natural echo path loss (or gain), degree of loss-switching, as well as level and/or type of singing (howling) protection are all important to a study of acoustic echo cancellers. In addition, it is important that an acoustic echo canceller is capable of working in harmony with a network-based electric echo canceller.

D.7.8 New circuit switched service

It has been suggested that there may be merit in modifying the disabling mode of G.165 cancellers so that upon the receipt of the disabling tone, the canceller disables until the connection is released.

It has been suggested that a customary procedure in some networks for initiating a digital transmission through a PCM-only digital voice network is to precede the digital transmission with a 2 100 Hz tone to disable any echo cancellers/suppressors in the circuit. However, the cancellers remain disabled only as long as the transmitted digital data, when interpreted as PCM samples, contain sufficient energy to maintain the cancellers in the disabled state. The success of this non-standard approach depends upon the content of the digital data stream, and, as the maintenance of a sufficient power level cannot be guaranteed, proprietary means are usually used to ensure that the cancellers remain disabled. When the disabling signal is digitally generated, additional complexity is required for terminals that use a bit-level protocol and a serial interface, due to the inability of the terminal to establish octet alignment with the octets used in the transmission channel.

In this context, the need for an in-band, non-octet aligned echo canceller disable signal is for further study. This study should be done in cooperation with T1S1 and CCITT SG I.

D.7.9 Comfort noise

As the telephone network migrates to more digital connections, it becomes more likely that the echo path will be analogue while the long distance connection path will be digital. One consequence is that the long distance path has a low idle channel noise while the echo path has a higher idle channel noise. This in turn leads to a situation called "noise modulation". When the NLP operates, the talker "hears" the idle channel noise of the digital long distance path, but when the NLP releases, the talker "hears" the idle channel noise of the echo path and the far-end environmental noise. Thus, the talker hears intervals of speech with background noise followed by intervals of silence, which can be very annoying in some instances.

There are two known approaches for comfort noise. The first solution is to insert pseudo-random noise during the silent interval. The second solution is to allow some of the background or idle channel noise to pass through the NLP. In the 1992 revision of CCITT Recommendation G.165 [19], the CCITT has provisionally accepted a test for comfort noise injection. This subject requires further consideration; the test, as proposed, is not accepted by some US manufacturers.

D.8 Special DCME/PCME networking considerations

It is well known that echo control is needed in long-delay circuits, such as on satellite links. In addition, echo control may be needed, even for a short terrestrial circuit, because of the additional buffering delay in a DCME or a PCME. If echo is present, it may be classified as speech and reduce the compression gain.

One possible interaction relates to the potential loading effect of the comfort noise injected by the echo canceller on a DCME/PCME (see figure D.4). The operation of the echo canceller may modulate the near-end analogue noise injected into the Sin port of the echo canceller. This could cause the adaptive speech detector of the DCME/PCME to falsely classify this change in noise level as the presence of speech. In this case, the DCME/PCME transmits the noise spurt as if it were speech and thus increases the activity factor of the circuit. The consequence is a decrease in the compression gain, and in some systems, an increase in the occurrence of freeze-out.

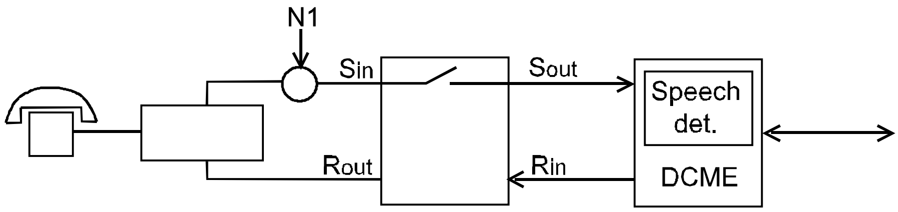


Figure D.4: Speech detector/Echo control device interactions

D.8.1 Detailed interaction

This interaction occurs as follows:

1) receive speech arrives at the receive input (R_{in}) port of the echo control unit;

2) the echo suppression switch or canceller NLP activates, stopping the echo or residual echo and attenuating the near-end-generated analogue terrestrial noise (N1) present at the send input (S_{in}) port;

3) if very little noise is generated between the echo control send output (S_{out}) port and the DCME speech detector input, the speech detector threshold adapts to its minimum level (typically, -50 dBm0);

4) when the receive speech stops, after a suitable echo control unit hangover time, the echo suppression switch or canceller NLP closes and the near-end-generated terrestrial noise (N1), as seen by the DCME speech detector, reappears as a step change in noise level;

5) the step change in noise level may exceed the speech detector threshold, causing the DCME to transmit a noise spurt as if it were speech. The noise spurt duration is a function of the adaptation speed of the speech detector and the near-end-generated terrestrial noise level.

This sequence is repeated for every speech spurt and produces a very annoying speech-correlated noise spurt heard by the far-end talkers every time they stop speaking.

This interaction is not limited to single echo control device network configurations. Figure D.5 shows a typical network configuration, with multiple echo control devices interacting with a DCME/PCME speech detector. In this configuration, the DCME/PCME speech detector may respond to unit step increases in noise power, which result from echo suppressor switch or echo canceller centre clipper activations in the send paths of echo control devices 1 and 3. (The role of the centre clipper is to remove the residual echoes due to imperfect cancellation). The DCME/PCME speech detector first experiences a unit step increase in noise power from echo control device 3 switch activation, followed by a second step increase from echo control device 1 switch activation. The extent to which the DCME/PCME speech detector incorrectly responds to these step increases in noise power is a function of the noise power levels N1, N2, N3, and N4, and the specific DCME speech detector threshold adaptation algorithm. For example, the dual step increases in noise presented to the DCME/PCME speech detector, which result from switch or centre clipper activation at locations 1 and 3, are masked if the power level N4 is excessively high. Likewise, high noise power levels at N2 or N3 may mask step increases in noise power caused by echo control unit 1.

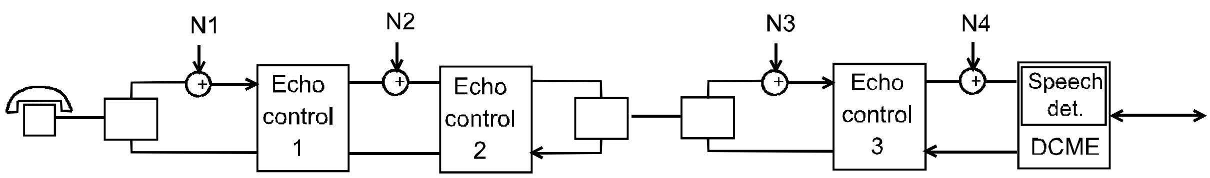


Figure D.5: Multiple echo control device configuration

D.8.2 Possible solutions

There are several methods for dealing with the interactions between the echo control devices and the DCME speech detector. In one approach, the echo control device could be modified to monitor the terrestrial-generated noise at the send-input port. When the send transmission path is broken, noise at the proper level is injected into the send-output toward the DCME, keeping the noise seen by the speech detector at a constant level (comfort noise) and avoiding speech detector activation. Not all echo cancellers may implement this approach, due to the number of different echo control devices in use and the uniqueness of this application.

In a second approach, the speech detector adaptive threshold of the DCME/PCME is frozen in the presence of speech on the corresponding receive channel.

A third approach is to specify an adaptive speech detector with a fast adaptation feature, which would track step changes in noise level and minimise the noise spurts.

The approaches described above may be unacceptable due to the number of different echo control devices in use and the uniqueness of the proposed application. Further, the large base of cancellers prevent consideration of a fast phasing in of new echo cancellers.

This subject requires further study and may result in recommendations to modify CCITT Recommendation G.165 [19] for new generation echo cancellers. The main point of this section is that the solution depends on the speech detection procedures of both the DCME/PCME and the echo canceller.

Annex E: A discussion of subjective tests

E.1 Introduction

The aim of this annex is to help the reader to *better understand the results* of subjective tests rather than to give instructions for how to perform them.

Much detailed information about subjective tests for telecommunication purposes can be found in the following ITU-T publications:

- Recommendation P.80 [85]: Methods for subjective determination of transmission quality;
- Recommendation P.83 [86]: Subjective performance assessment of telephoneband and wideband codecs;
- Recommendation P.84 [87]: Subjective listening test method for evaluating digital circuit multiplication and packetized voice systems;
- Handbook on Telephonometry [50].

Subjective tests should be carried out under "controlled conditions". Laboratory tests, conversational or listening, are designed to reproduce as far as possible the actual conditions met by the user. However, as is pointed out in the Handbook on telephonometry, these tests are by nature artificial and cannot provide all answers. A true measure of transmission quality under real conditions can only be achieved by customer surveys, for instance by interviewing customers after they have made actual calls. A suitable methodology is described in:

- Recommendation P.82 [67]: Method for evaluation of service from the standpoint of speech transmission quality.

Field surveys can give a valid score of the global transmission quality in a network but a disadvantage is of course the difficulty to ascertain the values of the transmission parameters which are involved in the calls observed.

Sometimes it is advisable to complement laboratory subjective tests of new equipment and/or systems with *field tests,* where prototype equipment is used by subjects who are encouraged to observe and note the performance of the system under a longer period of time and under realistic conditions.

E.2 Subjective evaluation criteria

The organisation and layout of a subjective test is very important and great care is necessary in order to get consistent results, not the least with regard to how the questions are put to the participants in the test. ITU-T Recommendation P.80 [85] recommends that proper instructions be prepared for the subjects, and that answers for judgements should follow certain patterns, chosen among the following layouts:

E.2.1 Absolute Category Ratings, ACR

Table E.1: Quality

Quality	Score
Excellent	5
Good	4
Fair	3
Poor	2
Bad	1

Table E.2: Listening effort

Listening effort	Score
Complete relaxation possible, no effort required	5
Attention necessary, no appreciable effort required	4
Moderate effort required	3
Considerable effort required	2
No meaning understood with any feasible effort	1

Table E.3: Loudness preference scale

Loudness preference scale	Score
Much louder than preferred	5
Louder than preferred	4
Preferred	3
Quieter than preferred	2
Much quieter than preferred	1

E.2.2 "General difficulty" - judgement

Additional useful information about speech transmission quality can often be obtained at the end of a subjective test by asking: "Did you experience any difficulty in communicating (listening); yes or no".

E.2.3 Quantal-Response Detectability Tests

Table E.4: Detectability opinion scale

Detectability opinion scale	Category
Objectionable	A
Detectable, but not objectionable	B
Not detectable	C

A, B and C should not be replaced by numbers for "averaging" purposes.

When a more detailed grading is needed, as for fading phenomena, etc., table E.5 is used.

Table E.5: Detectability opinion scale, more detailed grading

Detectability opinion scale	Category
Inaudible: noise completely undetectable	A
Just audible: noise can just be detected by listening carefully	B
Slight: noise detectable, but not disturbing	C
Moderate: noise slightly disturbing	D
Rather loud: noise causes appreciable disturbance	E
Loud: noise very disturbing, but call would be continued	F
Intolerable: noise so loud that the call would be abandoned	G

Here, A...G may be replaced by numbers, for instance 7...1, for computation of mean scores etc.

The actual conduct of these types of experiments resembles that of listening-effort tests but there are some differences, see ITU-Recommendation P.80 [85], Annex D. Where sidetone or echo is involved, the subject will be required to talk as well as listen.

E.2.4 Degradation Category Rating (DCR)

In this procedure for listening test, a quality reference condition is inserted before each judgement. (The choice of this reference depends on the application, i.e. the source signal is bandwidth limited to the same band as the application).

Table E.6: Degradation category

Degradation category	Score
Degradation is inaudible	5
Degradation is audible but not annoying	4
Degradation is slightly annoying	3
Degradation is annoying	2
Degradation is very annoying	1

E.2.5 The threshold method for comparison of transmission systems with a reference system

In this method, the system under test is assessed by comparison with a reference system having a variable degradation characteristic. See ITU-T Recommendation P.80 [85], annex E for details.

E.3 Statistical treatment and presentation of results, general

The raw data from subjective test requires in general a fair amount of statistical analysis in order to be interpreted properly. When the test layout is such that scores have been given numbers by the participants, the results are usually presented as numerical averages and designated

Mean Opinion Scores = MOS.

The MOS scale most often used is the one from 5 to 1.

When the same equipment or process is tested by different test teams, the spread in MOS values usually is in the order of 0,5 units for identical conditions, unless some strictly defined reference conditions are included.

Within the same experiment, a spread in MOS up to 0,7 can be observed.

Figure E.1 shows a typical spread in MOS judgements for the simple case of pure quantizing noise impairments from an MNRU, specified in ITU-T Recommendation P.81 [113], as evaluated by different test teams at different locations.

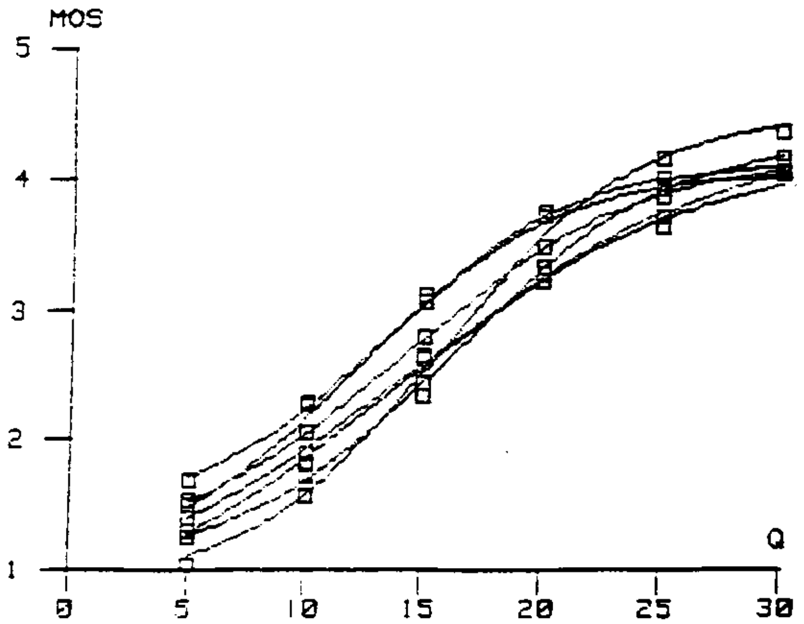


Figure E.1: MOS as function of quantizing distortion Q in a MNRU; seven different subjective tests

MOS figures are sometimes presented with two decimal places so one might be tempted to believe that a subjective test is a high-precision process, but this is definitely not the case. Therefore, before claiming that certain differences in MOS are "significant", a specific statistical analysis is needed.

Note that the choice of test arrangements and statistical evaluation methods depends very much on what answers the subjective test are expected to give, "selection" or "characterization". *Selection* implies finding the "best" among a number of candidate devices such as low bit-rate codecs. *Characterization* means describing the performance of a device, for instance with regard to talker dependency, effect of bit errors, environmental noise, transcoding with other devices etc., in terms which are meaningful for transmission planning.

In the selection type of work, by repeating experiments a sufficient number of times and applying sophisticated statistical methods, even "small" differences in MOS can be used as a basis for a "valid" choice of the best candidate.

For characterization tasks, one must keep in mind that the opinion of actual users can show a very large spread, a fact which cannot be altered by any mathematical treatment. Thus, there is a need to include certain safety margins for transmission planning when using results from this types of subjective tests. Also, repeating the subjective tests at different times may give results having a large spread from those originally used as a basis.

Finally, it might be worth pointing out the difference between *Reliability* and *Validity* as exemplified in figure E.2, showing targets at a shooting-range. In psychological test theory, it is usual to distinguish between reliability and validity. By reliability is meant how well a test measures what it does measure, while the validity tells how well a test measures what it should measure. In subjective tests, the *reliability* can be increased by using a large number of subjects and a sophisticated statistical analysis. To achieve a high *validity* the test layout must be worked out carefully to resemble real conditions. It is more important to have a high validity than a high reliability if a choice has to be made for any reason!

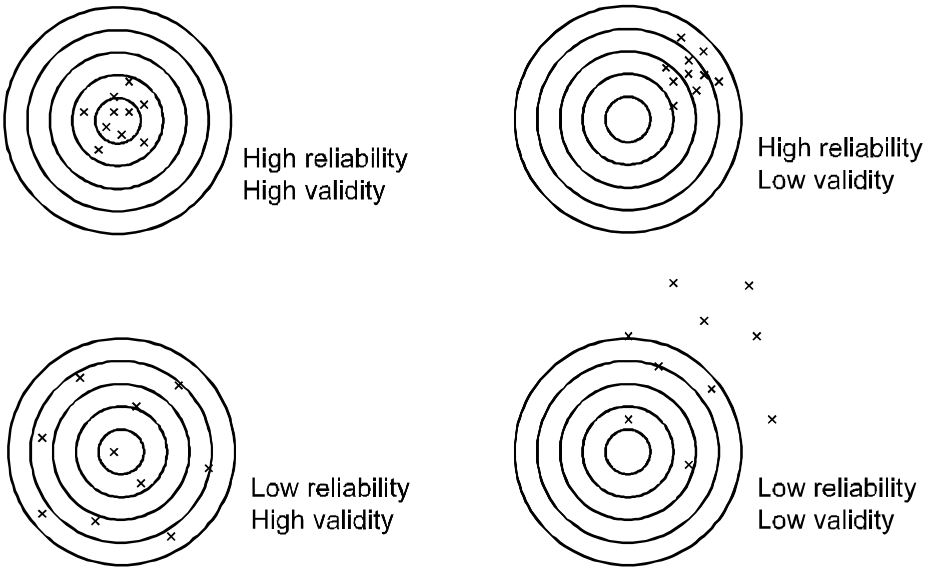


Figure E.2: Examples of *Reliability* and *Validity*

E.4 Handling of subjective test results

It is becoming increasingly more important to deal in a proper way with the subjective quality measures which are obtained in laboratory tests and from the field. This applies both for the analysis and the presentation of results because:

1) today and in the future, many new types of telecommunication networks and systems are being introduced. Therefore new, additional measures of quality will be needed;

2) there is a need for normalisation of such measures and for adopting procedures to increase their validity, accuracy and reliability;

3) it is desirable that common tools and format of data are being used in order to facilitate the analysis and comparisons of results obtained in different investigations.

E.4.1 Analysis of results

As mentioned before, the subjective test methods are mostly based on category judgement procedures which have been used for speech transmission quality assessment for a long time. However, other methods, based on questionnaire reports, paired comparisons, etc., have also been employed.

The results can be classified in a number of ways, a common one for subjective laboratory tests being MOS (Mean Opinion Score), as well as percentages of preferences. The results should be presented with values for confidence limits and standard deviations, using methods such as ANOVA (ANalysis Of VAriance) or ANCOVA (ANalysis of COVAriance). Other tools that may be applied are multidimensional analysis, factor analysis, general linear models, etc.

Some telecommunications administrations base their transmission planning on the percentage of GOOD+EXCELLENT (G+E) or POOR+BAD (P+B) ratings derived from a large amount of data obtained by the use of five-grade QUALITY scales (sometimes indicated as GOB (Good Or Better) and POW (Poor Or Worse)).

Recently an effort has been undertaken in the international research community in order to increase the reliability of results by special consideration of the number of responses and the test accuracy. As mentioned previously, MOS figures from a particular test are being transformed by using a reference curve into equivalent S/N ratios, given as equivalent Q dB values. The reference curve is obtained by means of a reference subjective test done in a set-up with a MNRU (Modulated Noise Reference Unit), using the same test team at the same occasion. This methodology has been widely adopted to compare the performance of different systems in terms of normalised units. More information about this methodology is given in annex G. (At present, work is going on to find an alternative reference impairment system which sounds more like the codecs being tested than the MNRU does).

In subclause E.4.3 a methodology is described for "normalisation" of MOS values from different test occasions by means of linear transformations.

Significance tests have been designed for specific experimental plans, taking into account the sensitivity of the customers to certain types of impairments such as delay or background noise or talker dependency. However, the advantage of getting an improved test precision should be weighted against additional time which is required to design very carefully defined test plans and analysis procedures.

E.4.2 Formats for the presentation of results

To evaluate properly the sometimes very complicated effects of transmission impairments, it is often necessary to evaluate and compare the results from different subjective tests.

Therefore graphs, histograms, figures, diagrams, scatter plots, should be produced in a "unified" way, i.e. by using the same magnitudes, metrics and units, so that they can be easily understood and correctly interpreted. This should especially be observed in internationally co-ordinated tests. Such tests are subjected to some inherent complications because they are conducted in different languages and employ people of different cultural backgrounds who may interpret the rating scales differently. In addition, sometimes competing business interests influence the test procedures and results.

On the other hand, this wish for "unification" should not exclude the use of recently developed powerful graphic tools which are better able to show the behaviour of complicated or composite factors. The use of spreadsheets, or of newly available calculation and visual tools for computer systems, including standard statistical analysis software, will make data formats from different organisations more easy to manage.

E.4.3 Comparison of MOS values, obtained from different test occasions or from model computations, by means of linear transformations

Subjective tests of impairment effects should be "normalised" by the inclusion of a "reference impairment" subjective test in order to "calibrate" the test team at the test occasion. However, most often when the same reference subjective test is made by different test teams and/or at different occasions, the reference MOS curves do not agree with any high degree of precision. (The same phenomena can be observed when a comparison is made with MOS from a computation model). Thus, there is a need to transform the different reference curves to a norm MOS curve in order to make appropriate comparisons between the actual subjective tests. A sophisticated method, used when evaluating low-bit-rate codecs, is described in annex G.

A simpler way to normalise is to use *linear transformations* between the different reference MOS curves and an agreed norm MOS curve, the linear transformation being determined by "anchor points". (In this context, the anchor points are so chosen that a "best curve fit" is obtained).

It can be assumed that an experiment with the impairment factor Q is performed and a curve with the relationship MOS = f_e (Q) is determined. This curve is to be fitted to a norm curve MOS = f_n(Q) with a simple linear expression.

To perform this, two anchor points are chosen where the transformed experimental curve and the norm curve are made to coincide, viz. Q_1 and Q_2 with the designations:

$$E_1 = f_e(Q_1);\; E_2 = f_e(Q_2);\; N_1 = f_n(Q_1);\; N_2 = f_e(Q_2)$$

Let

$$K = \frac{N_2 - N_1}{E_2 - E_1}$$

The expression for the transformed experimental curve MOS = f_t(Q) then becomes

$$f_t(Q) = f_n(Q_1) + K\left[f_e(Q) - f_e(Q_1)\right]$$

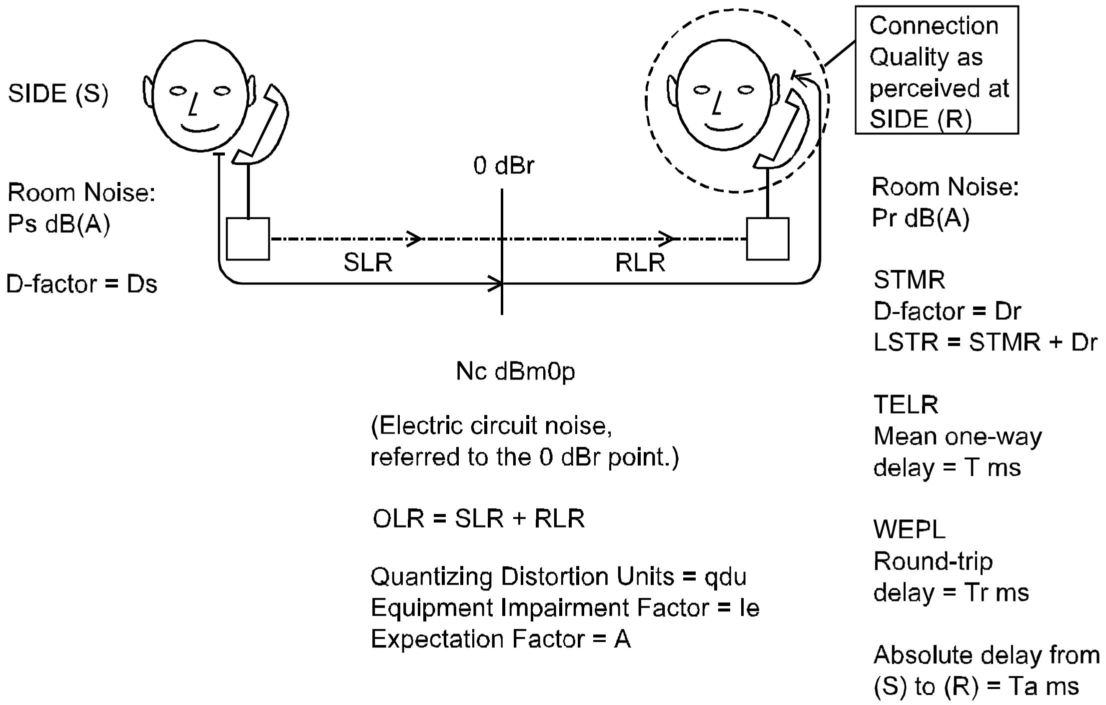
Annex F: Some examples of voice transmission quality evaluations by the ETSI model

F.1 General

The configuration of the connection and the relevant transmission parameters are as depicted in figure F.1 (the same as figure 38 in the main text). In what follows some examples will be given how certain parameters influence the voice transmission quality as perceived by a user in conversation mode, i.e. talking and listening over the connection.

When the quality as function of some parameter is illustrated by curves, the measure "Percentage of users finding the connection Good Or Better"(GOB) has been chosen as the Y-axis. (GOB indicates more clearly when the performance becomes questionable than for instance a MOS-value). Note that the values of those parameters, that are constant, are shown to the right of the curves.

When a particular connection is analysed the following voice transmission quality measures are given: R, GOB, POW (Poor Or Worse), TME (Terminate Early). (MOS figures have not been given because the typically large spread in MOS which may be observed between different subjective tests of the same phenomena).



NOTE: Evaluation of the voice transmission quality (R-factor, GOB, POW etc) refers to the user at side R, both as listener and talker.

Figure F.1: Configuration of the connection and the relevant transmission parameters

F.2 Loudness Ratings and electric circuit noise at the 0 dBr reference point

The example chosen is GOB as function of SLR with Nc (at the 0 dBr reference point) as a parameter (RLR = 3 dB; no other impairments in the connection). See figure F.2.

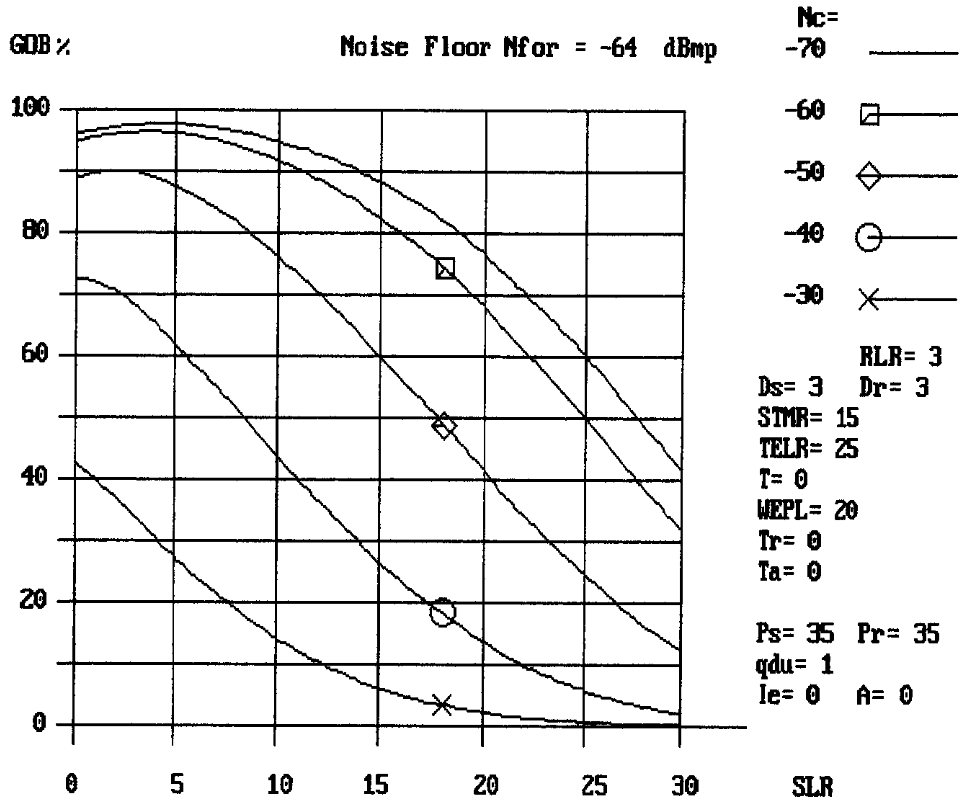


Figure F.2: GOB(SLR); Nc parameter

F.3 Influence of room noise at the send and receive sides

Figure F.3 illustrates the influence of room noise Pr dB(A) at the receiving side and the dependence on LSTR, in this case LSTR = STMR + Dr; Dr = 3 dB. (Dr = the D-factor of the set at the R-side).

Figure F.4 illustrates the influence of room noise Ps dB(A) at the send side and the dependence on Ds, the D-factor of the set at the send side.

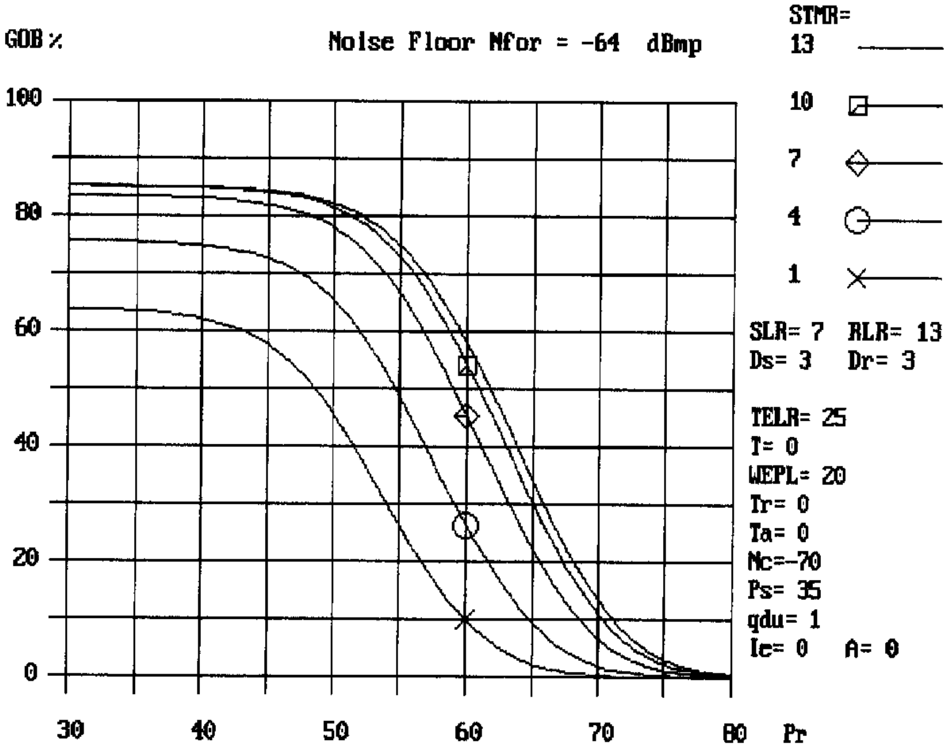


Figure F.3: GOB as function of room noise Pr dB(A) at the R-side

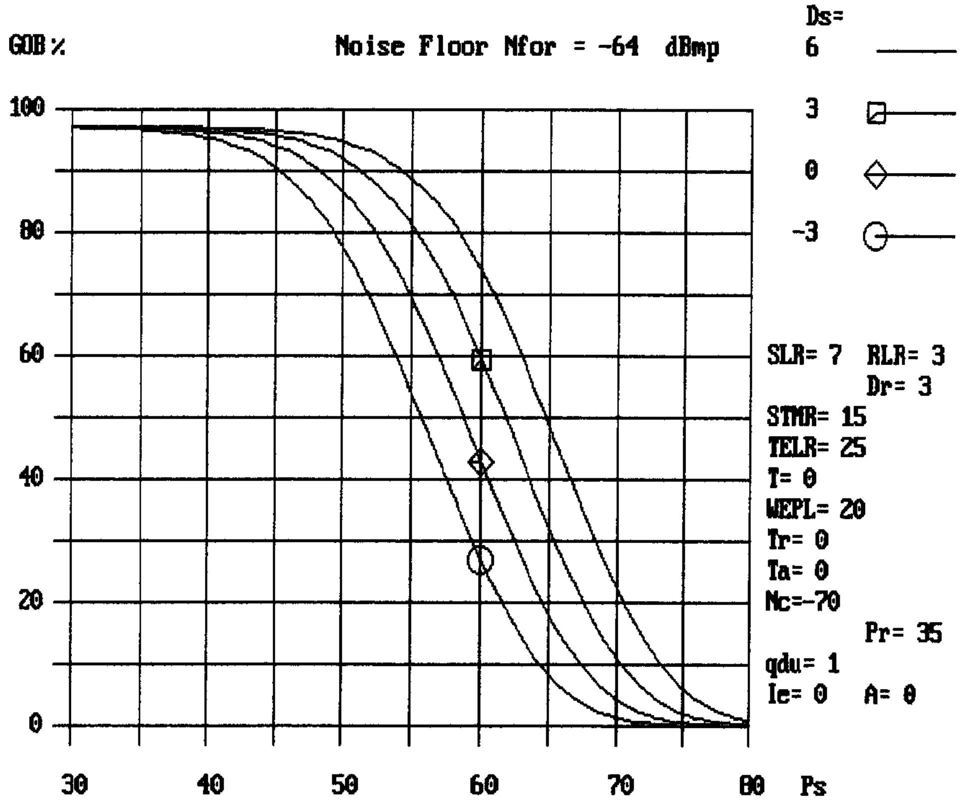


Figure F.4: GOB as function of room noise Ps dB(A) at the S-side

F.4 Listener echo, talker echo and long absolute delay

For listener echo, GOB as function of WEPL with the round-trip delay Tr ms as variable is shown in figure F.5. It appears that impairments due to WEPL will be rather minor in modern networks. The reason is that for longer delays the requirement on a sufficiently high TELR implies correspondingly high return losses at 2/4-wire hybrids so that WEPL > 25 or 30 dB.

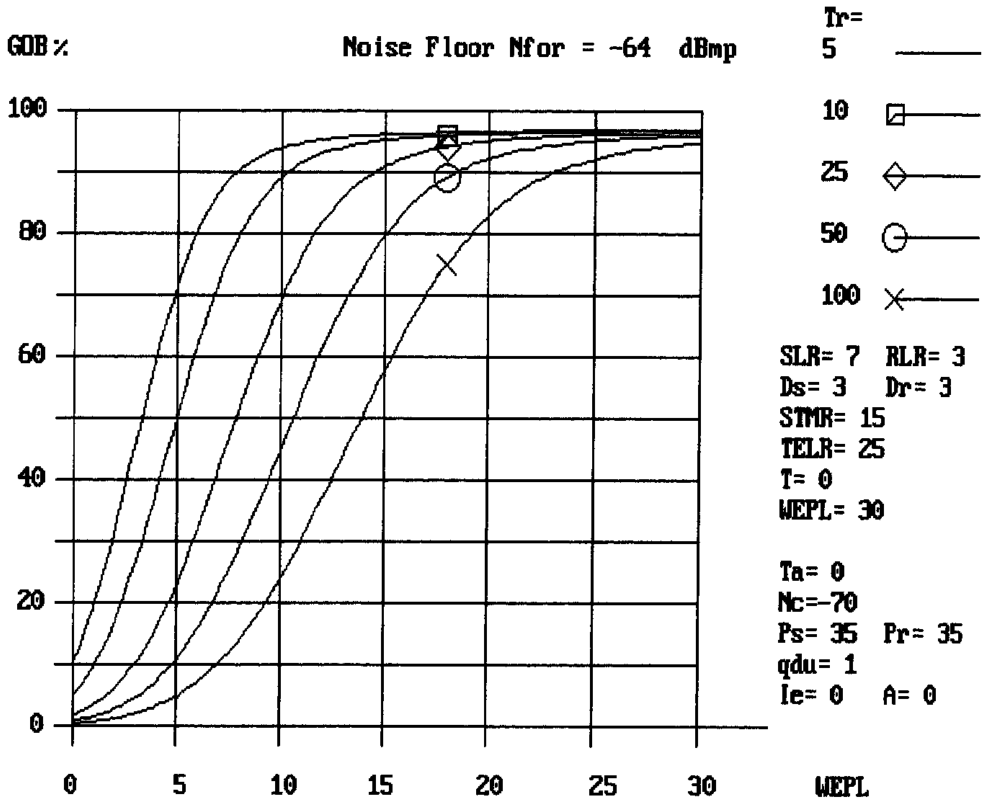


Figure F.5: GOB as function of listener echo WEPL dB; round-trip delay Tr ms as variable

There is a certain helpful masking effect of the talker sidetone on the talker echo. This is illustrated in figure F.6 for the case of TELR = 34 dB. As can be seen, in the absence of talker sidetone (STMR = 25 dB) the talker echo is more annoying than for the normal case (STMR = 15 dB). A very strong talker sidetone (STMR = 2 dB), on the other hand, has a certain masking effect on those strong talker echoes that occur for T > 40 ms. Some other cases of masking effects are shown in figure F.7.

For talker echo, GOB as function of mean one-way transmission time T ms with TELR as variable is shown in figures F.8 and F.9. In figure F.6 the time-scale is 1 - 100 ms, i.e. the effect of short-time echoes is included. In figure F.8 the time-scale is 10 - 300 ms and the effect of long absolute one-way transmission time Ta is included, the influence of which can be noticed above 200 ms. The latter effect is especially noticeable if the echo cancelling is less than perfect (low TELR).

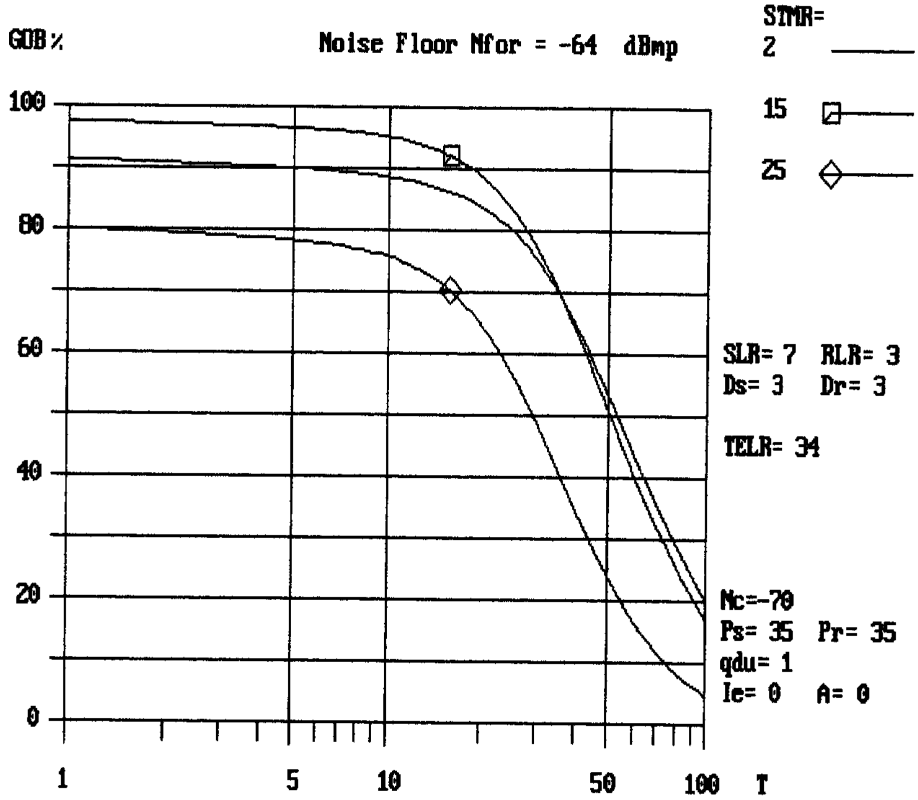
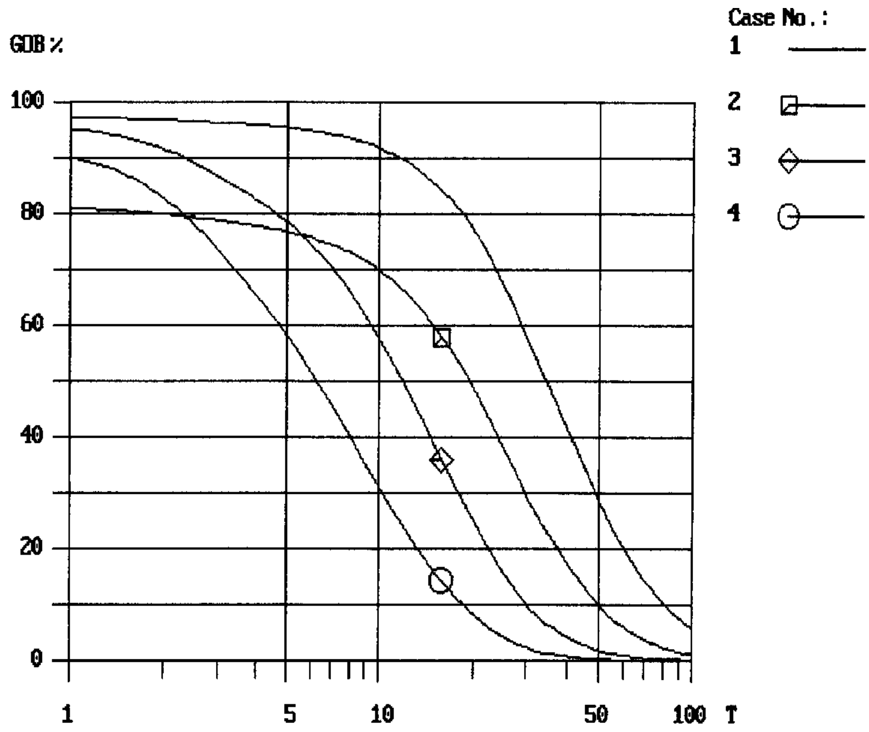


Figure F.6: GOB as function of mean one-way-delay for TELR = 34 dB; STMR as parameter



Case 1: STMR = 15 dB; TELR = 30 dB. Case 2: STMR = 25 dB; TELR = 30 dB. Case 3: STMR = 15 dB; TELR = 20 dB. Case 4: STMR = 25 dB; TELR = 20 dB

Figure F.7: GOB as function of mean one-way delay T ms illustrating the interaction between sidetone, normal resp. quiet, and echo. ("Normal" values of other parameters)

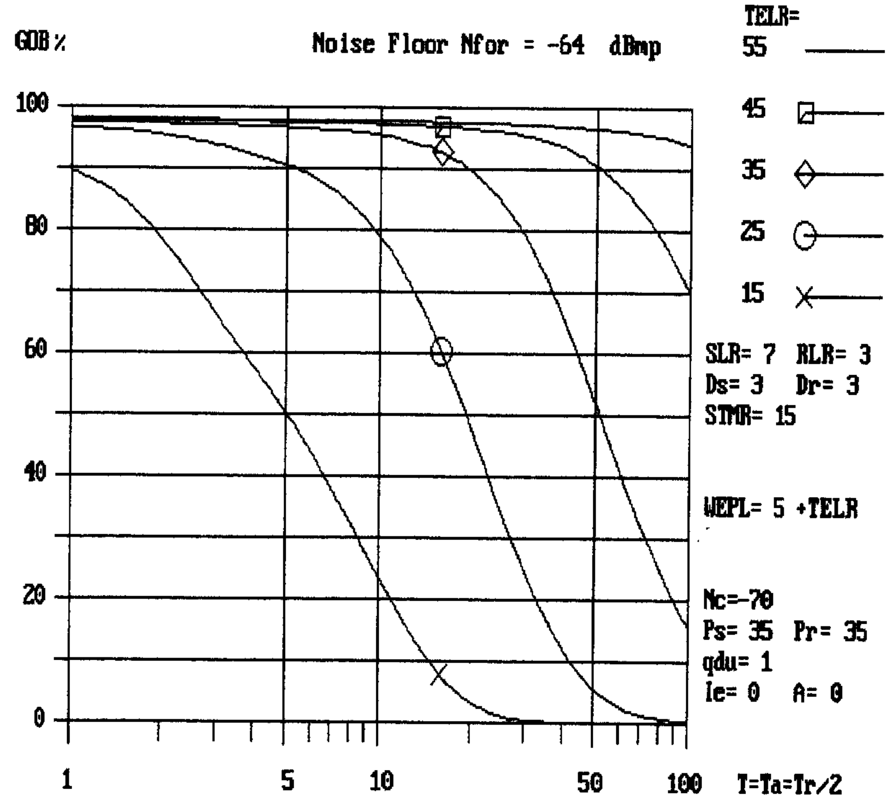
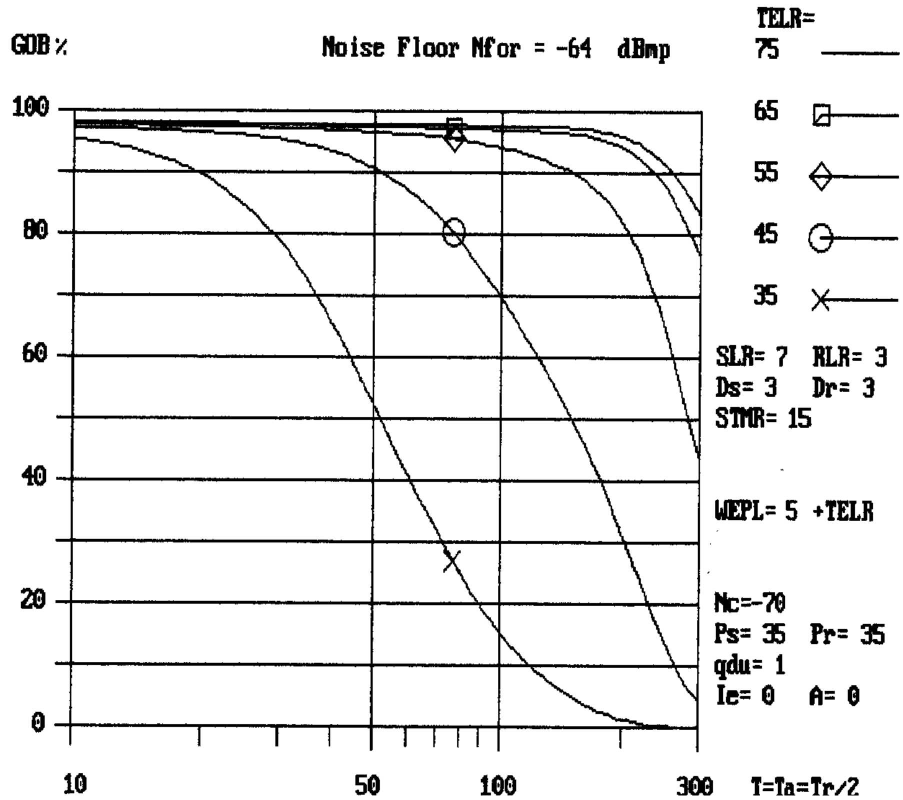


Figure F.8: The influence of talker echo. GOB as function of mean one-way delay T ms; TELR as variable

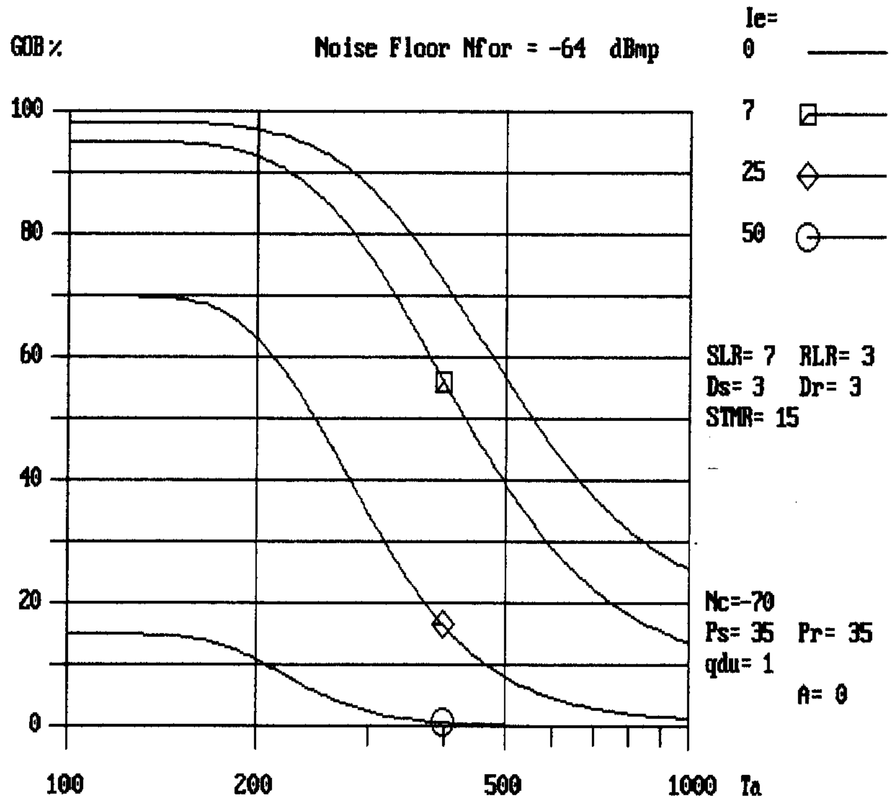


T = mean one-way delay for talker echo; Ta = one-way absolute delay from side S to side R; Tr =round-trip delay for listener echo, all measured in ms; TELR as variable.

Figure F.9: The influence of talker echo and long absolute delay on GOB

F.5 DCME using ADPCM 32, 24 and 16 kbit/s

Figure F.10 illustrates the effect of using a DCME (Digital Circuit Multiplication Equipment) in normal and compressed speech transmission modes, i.e. using ADPCM 32, 24 or 16 kbit/s, together with the effect of long one-way absolute transmission times Ta.



Ie = 7, 25, 50 corresponds to respectively ADPCM 32 kbit/s, 24 kbit/s, 16 kbit/s.

Figure F.10: The effect of DCME and long absolute one-way delay (Ta ms)

F.6 Some examples of connection configurations

NOTE 1: In the examples given, all transmission parameters for the model are tabulated in order to avoid misunderstandings. It is assumed that the listener echo is so weak that it can be neglected, as is the case for most modern networks. Thus the default values are set to WEPL = 20 dB, Tr = 0 ms.

NOTE 2: Even if the percentages GOB, POW etc. are presented with 0,1 "accuracy", this does not mean that one should expect this kind of agreement with actual customer opinion surveys.

F.6.1 Fully analogue connection

See figure 1, (subclause 4.2).

F.6.1.1 Normal

"Normal" with regard to SLR, RLR, Nc. Limiting case for no echo control. TELR = 35 dB, T = 25 ms.

SLR= 7 dB	RLR= 3 dB	OLR= 10 dB	Nc = -70 dBm0p
qdu= 1	Ie= 0	A= 0	(Nfor = -64 dBmp)
Abs. one way delay TA = 25 ms			
Side (S) :	Ds= 3 dB	Ps = 35 dB(A)	
Side (R)	Dr= 3 dB	Pr = 35 dB(A)	
		STMR= 15 dB	LSTR= 18 dB
	TELR= 35 dB	T (one-way) = 25 ms	
	WEPL= 20 dB	Tr (round-trip) = 0 ms	
R= 78,3			
GOB = 87,4 %	POW = 1,9 %	TME = 0,4 %	MOS = 3,96

F.6.1.2 Limiting loss

"Limiting loss", TELR increased due to the higher connection loss.

SLR= 15 dB	RLR= 6 dB	OLR= 21 dB	Nc = -60 dBm0p
qdu= 1	Ie= 0	A= 0	(Nfor = -64 dBmp)
Abs. one way delay TA = 25 ms			
Side (S) :	Ds= 3 dB	Ps = 35 dB(A)	
Side (R)	Dr= 3 dB	Pr = 35 dB(A)	
		STMR= 15 dB	LSTR= 18 dB
	TELR= 47 dB	T (one-way) = 25 ms	
	WEPL= 20 dB	Tr (round-trip) = 0 ms	
R= 72,6			
GOB = 78,4 %	POW = 4,2 %	TME = 1,1 %	MOS = 3,72

F.6.2 Digital - analogue connections

F.6.2.1 A fully digital connection with no special impairments

SLR= 7 dB	RLR= 3 dB	OLR= 10 dB	Nc = -70 dBm0p
qdu= 1	Ie= 0	A= 0	(Nfor = -64 dBmp)
Abs. one way delay TA = 0 ms			
Side (S) :	Ds= 3 dB	Ps = 35 dB(A)	
Side (R)	Dr= 3 dB	Pr = 35 dB(A)	
		STMR= 15 dB	LSTR= 18 dB
	TELR= 25 dB	T (one-way) = 0 ms	
	WEPL= 20 dB	Tr (round-trip) = 0 ms	
R= 93,6			
GOB = 98,2 %	POW = 0,1 %	TME = 0,0 %	MOS = 4,42

F.6.2.2 A digital/analogue connection with limiting delay

A digital/analogue connection with limiting delay for no echo cancellation but strong reflections at the analogue end; TELR = 25 dB, T = 25 ms.

SLR= 7 dB	RLR= 3 dB	OLR= 10 dB	Nc = -70 dBm0p
qdu= 1	Ie= 0	A= 0	(Nfor = -64 dBmp)
Abs. one way delay TA = -25 ms			
Side (S) :	Ds= 3 dB	Ps = 35 dB(A)	
Side (R)	Dr= 3 dB	Pr = 35 dB(A)	
		STMR= 15 dB	LSTR= 18 dB
	TELR= 25 dB	T (one-way) = 25 ms	
	WEPL= 20 dB	Tr (round-trip) = 0 ms	
R= 56,8			
GOB = 42,1 %	POW = 23,1 %	TME = 9,7 %	MOS= 2,93

F.6.2.3 A digital ATM connection

A digital ATM connection with terminating hybrid, see figure 2, (subclause 4.2). TELR = 35 dB, T = 25 + 9,5 = 34,5 dB.

SLR= 7 dB	RLR= 3 dB	OLR= 10 dB	Nc = -70 dBm0p
qdu= 1	Ie= 0	A= 0	(Nfor = -64 dBmp)
Abs. one way delay TA = 34,5 ms			
Side (S) :	Ds= 3 dB	Ps = 35 dB(A)	
Side (R)	Dr= 3 dB	Pr = 35 dB(A)	
		STMR= 15 dB	LSTR= 18 dB
	TELR= 35 dB	T (one-way) = 34,5 ms	
	WEPL= 20 dB	Tr (round-trip) = 0 ms	
R= 71,9			
GOB = 77,1 %	POW = 4,6 %	TME = 1,2 %	MOS = 3,68

F.6.2.4 A mobile (GSM) telephone connected to the PSTN

A mobile (GSM) telephone connected to the PSTN. "Perfect" echo cancellation.

SLR= 7 dB	RLR= 3 dB	OLR= 10 dB	Nc = -70 dBm0p
qdu= 1	Ie= 20	A= 10	(Nfor = -64 dBmp)
Abs. one way delay TA = 95 ms			
Side (S) :	Ds= 3 dB	Ps = 35 dB(A)	
Side (R)	Dr= 3 dB	Pr = 35 dB(A)	
		STMR= 15 dB	LSTR= 18 dB
	TELR= 75 dB	T (one-way) = 95 ms	
	WEPL= 20 dB	Tr (round-trip)= 0 ms	
R= 82,9			
GOB = 92,4 %	POW = 0,9 %	TME = 0,2 %	MOS = 4,13

F.6.2.5 A mobile (GSM) telephone connected to the PSTN via a satellite link

A mobile (GSM) telephone connected to the PSTN via a satellite link, including a DCME. See figure 3, (subclause 4.2).

SLR= 7 dB	RLR= 3 dB	OLR= 10 dB	Nc = -70 dBm0p
qdu= 1	Ie= 27	A= 10	(Nfor = -64 dBmp)
Abs. one way delay TA = 395 ms			
Side (S) :	Ds= 3 dB	Ps = 35 dB(A)	
Side (R)	Dr= 3 dB	Pr = 35 dB(A)	
		STMR= 15 dB	LSTR= 18 dB
	TELR= 75 dB	T (one-way) = 395 ms	
	WEPL= 20 dB	Tr (round-trip) = 0 ms	
R= 51,1			
GOB = 28,9 %	POW = 35,2 %	TME = 17,3 %	MOS = 2,63

F.6.2.6 DECT internal system

DECT internal system. Talker echo caused by the acoustic echo; TELR = 50 dB.

SLR= 7 dB	RLR= 3 dB	OLR= 10 dB	Nc = -70 dBm0p
qdu= 1	Ie= 14	A= 5	(Nfor = -64 dBmp)
Abs. one way delay TA = 28 ms			
Side (S) :	Ds= 3 dB	Ps = 35 dB(A)	
Side (R)	Dr= 3 dB	Pr = 35 dB(A)	
		STMR= 15 dB	LSTR= 18 dB
	TELR= 50 dB	T (one-way) = 28 ms	
	WEPL= 20 dB	Tr (round-trip) = 0 ms	
R= 82,0			
GOB = 91,6 %	POW = 1,0 %	TME = 0,2 %	MOS = 4,10

F.6.2.7 DECT in a private network

DECT in a private network with 3 DCME in tandem, "perfect" echo cancelling.

SLR= 7 dB	RLR= 3 dB	OLR= 10 dB	Nc = -70 dBm0p
qdu= 1	Ie= 28	A= 5	(Nfor = -64 dBmp)
Abs. one way delay TA = 75 ms			
Side (S) :	Ds= 3 dB	Ps = 35 dB(A)	
Side (R)	Dr= 3 dB	Pr = 35 dB(A)	
		STMR= 15 dB	LSTR= 18 dB
	TELR= 75 dB	T (one-way) = 75 ms	
	WEPL= 20 dB	Tr (round-trip) = 0 ms	
R= 70,0			
GOB = 73,4 %	POW = 5,9 %	TME = 1,7 %	MOS = 3,60

F.6.2.8 Private network

Private network using a fictitious, non-standard codec with Ie = 30, codec-delay = 15 ms one-way. Network one-way delay = 25 ms, digital telephones with TCLw = 40 dB, i.e. TELR = 50 dB.

SLR= 7 dB	RLR= 3 dB	OLR= 10 dB	Nc = -70 dBm0p
qdu= 1	Ie= 30	A= 0	(Nfor = -64 dBmp)
Abs. one way delay TA = 40 ms			
Side (S) :	Ds= 3 dB	Ps = 35 dB(A)	
Side (R)	Dr= 3 dB	Pr = 35 dB(A)	
		STMR= 15 dB	LSTR= 18 dB
	TELR= 50 dB	T (one-way) = 40 ms	
	WEPL= 20 dB	Tr (round-trip) = 0 ms	
R= 59,7			
GOB = 49,1 %	POW = 18,0 %	TME = 7,0 %	MOS = 3,08

Annex G: Consideration of low bit-rate codecs in transmission planning

G.1 Introduction

Sometimes today, and often in the future, telephone networks subscribers will encounter a connection with one or more low bit-rate codecs. A careful transmission planner therefore needs some planning rules in this respect. The "qdu methodology" as given in Recommendation G.113 [35] appears not to be quite adequate for low bit-rate codecs. In the ETSI computation model, another method is proposed which is based on "equipment impairment factors". The actual values proposed for the equipment impairment factors, here designated *Ie*, are derived from a number of subjective tests recently published. (An advantage of the "impairment factor" methodology is that the effect of the combination of several different impairments can be evaluated).

G.2 Subjective tests, Mean Opinion Scores (MOS) evaluations

For obvious reasons, opinion scores for the speech quality of low bit-rate codec have a rather low degree of precision. The codec speech impairments are of many different kinds and they do not sound at all in the same way as pure circuit noise and quantizing distortion. Participants in subjective experiments do not have clear references to comparable, "quantified" impairments. It may well happen that codecs are ranked in different order with regard to speech quality in experiments made at different times and in different laboratories.

MOS values in subjective tests are sometimes presented with two decimal places so one might think that determining Mean Opinion Scores is a high-precision process. However, when different laboratories make the same subjective test of a particular process, the MOS values obtained may differ considerably. This is further discussed in annex E.

To compare two subjective tests (in the form of MOS values 1 - 5) one can make a linear transformation of the values from one set to the other by using "anchor points" for comparable conditions at either end of the MOS scale, for instance for upper and lower MOS saturation values. This methodology is described in annex E, subclause E.4.3.

Another, more sophisticated method is to use "equivalent Q" values obtained by including in the subjective test some reference conditions with quantizing distortion from a MNRU (Modulated Noise Reference Unit).This is described in reference [2] and subclause G.3.

Both methodologies have been used when the codec equipment impairment factors were derived for the ETSI model. The procedure is described in reference [4] and subjective test results from references [2], [11], [57], [70] and [71] were used. At present, the following codecs are included in the ETSI model:

- ADPCM 40, 32, 24, 16 kbit/s;
- LD-CELP 16, 12,8 kbit/s;
- VSELP 8 kbit/s;
- RPE-LTP 13 kbit/s (i.e. the GSM codec);
- CELP+ 6,8 kbit/s.

Discussion:

References [11] and [70] use a common database for subjective tests. Reference [70] includes tests with VSELP and reference [11] tests with LD-CELP 12,8 kbit/s.

The "saturation MOS value" for a very good connection with only one 64 kbit/s PCM link appears to be MOS = 4,3 in references [11] and [70]. (The same saturation value is MOS = 4,3 in the ETSI model and MOS = 4,8 in the TQI model in Annex A to ITU-T Recommendation P.11 [26]).

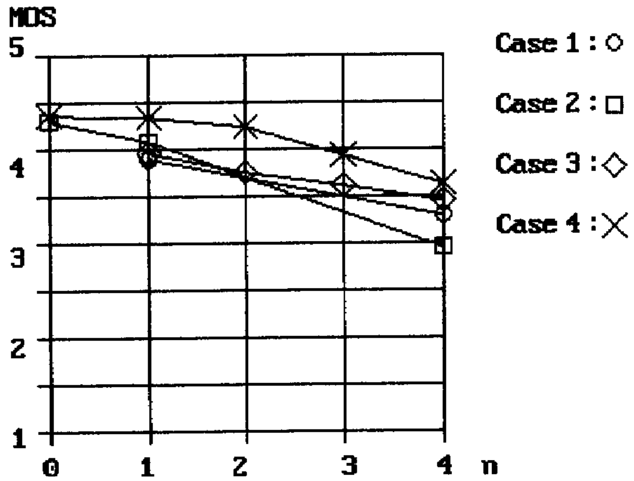
As expected, a certain spread in MOS values can be noted between results from the different references, especially for codecs with bit-rates lower than 16 kbit/s. (Up to 0,5 for ADPCM 32 kbit/s, 0,7 for VSELP 8 kbit/s).

Even in the simple case of pure quantizing noise an appreciable spread in MOS can be observed between different subjective tests. Figure E.1 shows curves of MOS as a function of MNRU quantizing distortion Q as they are given in references [2], [11], [57], [70] and [71].

As found in reference [4], the MOS values decrease approximately linearly with the number of codecs in tandem. This applies also for ADPCM 32 kbit/s.

Regarding ADPCM 32 kbit/s an interesting comparison can be made with the traditional "qdu" planning rule. The "qdu method" implies that the quantizing distortion adds to the circuit noise so that the resulting noise acts like an increased circuit noise, thereby affecting the voice transmission quality. For low qdu values the normal circuit noise dominates. Thus, the decrease in MOS should be very slight when the qdu value is increased gradually from 0.

The qdu methodology is also implemented in the ETSI model. One ADPCM 32 kbit/s codec is usually considered to have qdu = 3,5 and in figure G.1 the corresponding computed MOS curve, marked "X", is drawn. As can be seen, the discrepancies between this and the actual MOS curves are rather large, indicating that the quantizing distortion from ADPCM does not influence the quality in direct proportion to the corresponding increase in equivalent circuit noise but rather more, at least for up to 3 tandem-connected codecs. Therefore, the qdu method may not be quite satisfactory for ADPCM 32 kbit/s.



Case 1: Reference [56]. Case 2: reference [70]. Case 3: reference [7]. Case 4: Predictions by using qdu = 3,5.

Figure G.1: ADPCM 32 kbit/s; MOS as function of n codecs in tandem

G.3 Derivation of the equipment impairment factors *Ie*, general principles

To derive the impairment factor for a particular equipment, like a combination of low bit-rate codecs, from subjective measurements the following procedures are adopted:

The most straightforward method is a direct consideration of MOS figures obtained from a particular experiment. First an "anchor point" with an "upper saturation value" MOS = MOSu is established for the case when the connection does *not* contain the equipment under study, i.e. a "clear channel" with a very small speech degradation, like a "normal" PCM channel. In the ETSI model, the "clear channel" has a (computed) value MOS(E)u = 4,33. (SLR = 7 dB, RLR = 3 dB; other impairments are negligible).

The "lower saturation value" MOSl is MOS(E)l = 1 for the ETSI model. Often, but not always, also in subjective tests MOS(S)l = 1.

Thus, one can make a linear transformation between the "subjective" MOS and the "model" MOS as described in annex E, subclause E.4.3.

A more sophisticated method to evaluate codec test results from several different experiments is to normalise the MOS figures by the so-called "equivalent Q" method. This is based on comparisons with a MNRU-created quantizing distortion Q dB. (MNRU = Modulated Noise Reference Unit).

In the "equivalent Q" method the relation between MOS and Q can be shown to have, with good accuracy, the following general relation:

$$MOS = 1 + A + B \cdot \tanh\left[\frac{Q - Q_m}{C}\right] \qquad \text{(G.1)}$$

where *A, B, C* and *Qm* are constants. The test team gives MOS figures for a number of quantizing distortions measured objectively by Q and the constants are chosen so as to give the best fit between the subjective and the calculated MOS for the MNRU tests. The MOS figure for the *codec under test* is then transformed into an "equivalent Q"-value

$$Q = Q_m + \frac{C}{2} \cdot \ln \frac{B - A - 1 + MOS}{B + A + 1 - MOS} \qquad \text{(G.2)}$$

This latter method has often been used to present results from tests of low bit-rate codecs at different laboratories in order to allow for easy comparisons. (Note, however, that low bit-rate codec impairments do not sound as quantizing distortion and this limits the validity).

The ETSI model allows the computation of MOS as function of qdu (quantizing distortion units) with the commonly used relation

$$Q = 37 - 15 \cdot \lg(qdu) \qquad \text{(G.3)}$$

and the constants in equ.(G.1) turn out to be, for a very good fit for 5 < Q <26,

$$A = 2{,}20;\ B = 1{,}48;\ C = 7{,}84;\ Q_m = 17{,}2 \qquad \text{(G.4)}$$

Thus, it is possible to get the ETSI model MOS-values also from the equivalent Q values.

If the input data are in the form of *subjective* MOS-values a linear transformation according to annex E, subclause E.4.3 gives the ETSI MOS(E) and by means of the relation

$$MOS = 1 + 0{,}035 \cdot R + R(R - 60)(100 - R) \cdot 7 \cdot 10^{-6} \qquad \text{(G.5)}$$

a value of *R(E)* can be obtained (by interpolation).

The "clear channel" MOS(E) = 4,3 corresponds to *R =* 90. Thus, the "equipment impairment factor" is

$$Ie = 90 - R(E) \qquad \text{(G.6)}$$

If the input is in the form of *equivalent* Q-values, the corresponding impairment factor can be obtained directly, see subclause 9.1.3.3:

$$Ie = Iq = 15 \cdot \lg\left[1 + 10^{(Ro - 100)/15} \cdot 10^{(46 - G)/10}\right] \qquad \text{(G.7)}$$

where *Ro* = 95 and

$$\mathrm{G} = 1{,}07 + 0{,}258.\mathrm{Q} + 0{,}0\,602.\mathrm{Q}^2 \qquad \text{(G.8)}$$

Vice versa, the value of Q can also be computed from a known value of the impairment factor *Ie.* First *G* is computed

$$G = (Ro - 31)/1{,}5 - 10 \cdot \lg\left[10^{Ie/15} - 1\right] \qquad \text{(G.9)}$$

The value of Q is then obtained by solving the simple second-degree equation (G.8).

Note that the relation between MOS(E) and R is almost linear in the range 1,5 < MOS(E) < 4,5 with a slope

$$\frac{dR}{d(MOS)} \approx 20 \qquad \text{(G.10)}$$

G.4 Derivation of the equipment impairment factors *Ie* from the subjective test data

NOTE: Using several different subjective tests may increase the *reliability* but in general not the *precision* of the results. Due to the imprecise nature of MOS figures, it is to be expected that the derived *Ie* factors also can be similarly imprecise. However, for *transmission planning* the important thing is to *agree* upon *Ie* values which will provide reasonable estimations, with certain safety margins, of how the users will judge the voice transmission quality in a connection as it is affected by the equipment in question.

To illustrate the methodology for deriving codec impairment factors, the procedure used in reference [4] will be described.

From the results presented in reference [4] it was clear that the MOS values approximately *decrease* in direct proportion to the number of (like) codecs in tandem. Due to the almost linear relation between MOS and R, it could be expected that *Ie increases* linearly with the number of like codecs in tandem.

The first step was to investigate the cases with codecs *of the same type* in tandem and deriving the corresponding *Ie* values for each type of codec. The second step was to see if the subjective MOS values for a tandem connection of a mixture of codecs *of different types* could be predicted by adding the *Ie* values of the individual codecs in the chain.

Values of MOS(E) have been derived from the data in references [2], [11], [70] and [71] by means of the "equivalent Q" values and from data in reference [57] by transforming the subjective MOS values with the help of anchor points. The parameters *A...Qm* in equ.(G.3) and (G.4), applicable for references [2], [70], [11] and [71] respectively, are given in table G.3. The MOS(E) values have then been converted to *Ie* values by means of equ.(G.5) and (G.6) or (G.7). As an example, table G.1 presents the results for ADPCM 32 kbit/s and LD-CELP 16 kbit/s.

Table G.1: Equipment impairment factor *Ie* for *n* codecs in tandem

Codec	*n*	Reference [1]	Reference [2]	Reference [3]	Reference [4]
ADPCM 32	1	3,1	5,5	3,7	8,0
	2	12,3	-	-	18,5
	3	-	-	-	23,4
	4	23,3	-	27,7	28,4
LD-CELP 16	1	1,7	-	6,9	7,7
	2	11,5	-	-	17,4
	3	14,3	-	19,9	23,4
	4	22,1			30,5
	4	22,1	-	-	30,4

The *Ie*-values in table G.1 can be approximated by the simple relation

$$Ie = n \cdot K \qquad \text{(G.11)}$$

where K = 6,8 for ADPCM 32 kbit/s and K = 6,5 for LD-CELP 16.

The standard deviations between the values in the table and the approximations are

St.D. = 2,9 for ADPCM, St.D. = 3,7 for LD-CELP 16.

corresponding to a variation in MOS of about 0,2.

For transmission planning one can put K = 7 for both ADPCM 32 and LD-CELP 16. (G.12)

Further numeric analysis of data for the other types of codecs showed that the linear relation in equ.(G.11) also could be used for those.

Table G.2 gives results for ADPCM (G.726) of varying bit-rates.

Table G.2: MOS(S) and equipment impairment factor *Ie* for ADPCM (G.726), varying bit-rate; data from references [57] and [70]

kbit/s	MOS(2)	MOS(3)	*Ie(2)*	*Ie(3)*
(PCM 64)	3,95	4,30	0	0
40	3,91	-	1,7	-
32	3,81	(4,09)	5,5	7
24	3,46	3,23	15,9	23,3
16	2,61	1,99	35,6	41,9

As can be seen, the subjective scores differ considerably between the two investigations and as a consequence of this, also the derived impairment factors. For transmission planning purposes it seems appropriate to use the more pessimistic results of reference [71] as a guidance with the values

For ADPCM 40 kbit/s K = 2;
for ADPCM 24 kbit/s K = 25; (G.13)
for ADPCM 16 kbit/s K = 45.

Similar computations have been made for the other types of codecs. In particular, using data from references [70] and [71]:

For VSELP 8 kbit/s K = 16 (G.14a)

However, using data *only* from reference [71] the result was:

For VSELP 8 kbit/s K = 19 (G.14b)

Using data from reference [11]:

For LD-CELP 12,8 kbit/s K = 20 (G.15)

Using data from reference [71]:

For RPE-LTP 13 kbit/s K = 19 (G.16)

For CELP+ 8 kbit/s K = 22 (G.17)

Table G.3: MOS curve parameters for equ.(G.1)

Used in	*A*	*B*	*C*	*Qm*
Reference [2]; Table 18	1,64	1,56	10,1	14
Reference [70]; Table 1	1,69	1,83	8,18	15,7
Reference [11]; Table 1	1,69	1,83	8,18	15,7
Reference [71]; Table 5-6	1,519	1,552	8,43	11,35

G.5 Some comparisons between measured and predicted MOS values

References [2], [11], [70] and [71] present subjective tests in the form of MOS(S) figures for various combinations of codecs in tandem. Using equ.(G.13) - (G.17), predicted mean opinion scores according to the ETSI model have been computed and compared with the subjective ones. The results are given in tables G.4 - G.7 and depicted in figure G.2.

(For reference [71], only its table 5-6 has been used because this is the most comprehensive one).

Note that in these comparisons the actual subjective MOS values obtained at each particular test occasion were used. The predicted MOS values were computed by the transformation from predicted equivalent Q-values, using those curve parameters which belonged to the corresponding test occasion. In this way, the variations in the subjective MOS perception between the test occasions were eliminated from the comparison results. (In normal use of the computation model for evaluation of a network configuration, no such transformation of MOS values would be used).

Table G.4: Comparison between subjective MOS(S) and computed predicted MOS(P)
Reference [32] table 18

Codecs	MOS(S)	MOS(P)	MOS(S) - MOS(P)
LD	3,93	4,14	-0,21
AD	3,88	4,14	-0,26
LD x 2	3,67	3,88	-0,21
AD x 2	3,65	3,88	-0,23
LD + AD + LD	3,61	3,57	0,04
LD x 3	3,60	3,57	0,03
AD + LD + AD	3,52	3,57	-0,05
LD x 4	3,38	3,22	0,16
AD x 4	3,34	3,22	0,12
NOTE: ADPCM 32 kbit/s = AD, LD-CELP 16 kbit/s = LD. (K = 7).			

Table G.5: Comparison between subjective MOS(S) and computed predicted MOS(P)
Reference [70], table 1

Codecs	MOS(S)	MOS(P)	MOS(S) - MOS(P)
PCM (64)	4,30	4,35	-0,05
AD 32	4,09	4,14	-0,05
LD	3,96	4,14	-0,11
VS	3,85	3,66	0,19
VS + LD	3,51	3,32	0,19
LD + LD + LD	3,41	3,57	-0,16
LD + VS + LD	3,26	2,96	0,30
AD 24	3,23	3,37	-0,14
AD 32 + LD + AD 32	3,16	3,57	-0,41
VS + VS	3,13	2,70	0,43
VS + AD 32 + LD	3,10	2,96	0,14
4 x (AD 32)	2,97	3,22	-0,25
VS + LD + VS	2,69	2,33	0,36
AD 16	1,99	2,33	-0,34
NOTE: ADPCM 32, 24, 16 kbit/s = AD 32, 24, 16. (K = 7, 25, 45) LD-CELP 16 kbit/s = LD. (K = 7) VSELP 8 kbit/s = VS. (K = 19).			

**Table G.6: Comparison between subjective MOS(S) and computed predicted MOS(P)
Reference [11], table 1**

Codecs	MOS(S)	MOS(P)	MOS(S) - MOS(P)
PCM (64)	4,30	4,35	-0,05
LD 12	3,14	3,61	-0,47
LD 16 + LD 12	3,11	3,27	-0,16
LD 16 + LD 12 + LD 16	2,84	2,91	-0,07
AD + LD 12 + AD	2,82	2,91	-0,09
LD 16 + LD 12 + AD	2,67	2,91	-0,24
LD 12 + LD12	2,60	2,59	0,01
LD 12 + LD 16 + LD 12	2,38	2,23	0,15
3 x (LD 12)	2,13	1,62	0,51
NOTE: ADPCM 32 kbit/s = AD. (K = 7). LD-CELP 16 kbit/s = LD 16. (K = 7). LD-CELP 12,8 kbit/s = LD 12. (K = 20).			

**Table G.7: Comparison between subjective MOS(S) and predicted MOS(P)
Reference [71], table 5-6**

Codecs	MOS(S)	MOS(P)	MOS(S) - MOS(P)
LD	3,96	4,14	-0,18
AD	3,95	4,14	-0,19
LD x 2	3,79	3,88	-0,09
AD x 2	3,76	3,88	-0,12
LD x 3	3,63	3,57	0,06
AD x 3	3,62	3,57	0,05
RPE-LTP	3,60	3,66	-0,06
CELP+	3,60	3,52	0,08
VS	3,54	3,66	-0,12
AD x 4	3,46	3,22	0,24
LD x 4	3,38	3,22	0,16
RPE-LTP x 2	3,09	3,22	-0,13
CELP+ x 2	2,69	2,39	0,30
VS x 2	2,52	2,70	-0,18
NOTE: VSELP 8 kbit/s = VS. (K = 19) ADPCM 32 kbit/s = AD. (K = 7) LD-CELP 16 kbit/s = LD. (K = 7) RPE-LTP 13 kbit/s = RPE-LTP. (K = 19) CELP+ 8 kbit/s = CELP+. (K = 22).			

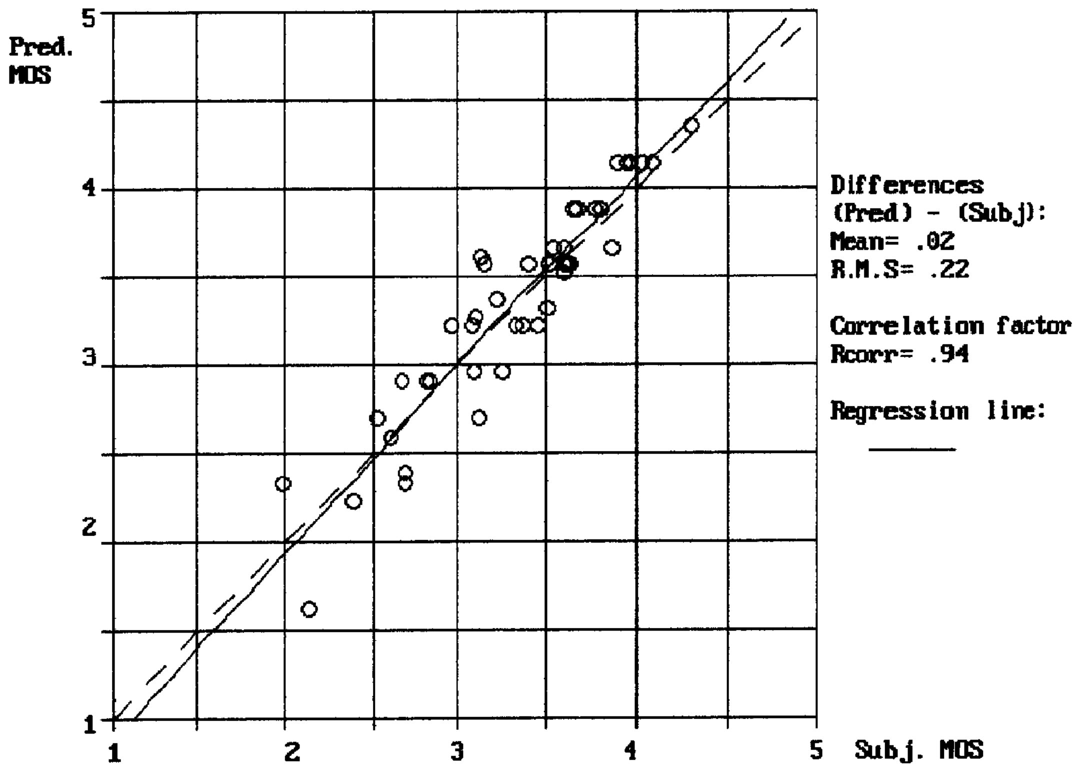


Figure G.2: Relation between subjective and predicted MOS as given in tables G.4 - G.7

The agreement between subjective and predicted MOS values is in general as close as is usually the case between two well-controlled subjective experiments and must be considered as quite adequate for transmission planning. (Note that the predicted MOS values were derived directly from the ETSI model without any adjustment with regard to the individual MOS "equivalent Q" reference curves obtained in the individual subjective tests. It appears that the ETSI MOS function represented a fairly good average of the subjective tests).

Also note that a MNRU is not a very good reference for the type of speech impairments caused by low bit-rate codecs which fact decreases the validity of MOS. One can note the fairly large spread between equivalent Q-values found in references [11], [70] on one hand and in reference [71] on the other.

MOS figures relate in principle to a particular set of subjective tests under controlled conditions and for a reliable interpretation some form of anchoring to reference conditions should be included. However, for practical use in transmission planning, MOS figures as such are less suitable. Instead, "nominal customer reactions" in the form of percentages of "GOB", "POW" and/or "TME" are to be preferred as indicators of the speech transmission quality in a network. (GOB = good or better, POW = poor or worse, TME = terminate early).

For use in transmission planning some slightly modified values of *K* will be introduced in subclause G.7 in order to provide for more safety margins in evaluations of speech transmission quality. Note the typical spread in peoples' opinion of speech transmission quality as it is depicted in figure E.1. (In transmission planning one cannot use individual "anchor points" each time a customer survey is made in order to normalize the subjective opinions).

G.6 Comparison between the "quantizing distortion unit" qdu and the "equipment impairment factor *Ie*

From what has been shown in the previous sections, it appears that the "equipment impairment factor" *Ie* for low bit-rate codec has additive properties. It is interesting to compare it with the "quantizing distortion unit" qdu which is considered additive for normal PCM.

Figure G.3 shows both *Ie* and qdu as functions of the MNRU quantizing distortion Q. As can be seen, there is quite a large difference between the two quantities which may explain why the qdu concept is less applicable for *low bit-rate codecs.* For those, adding qdu values will not give very relevant results with regard to speech transmission quality.

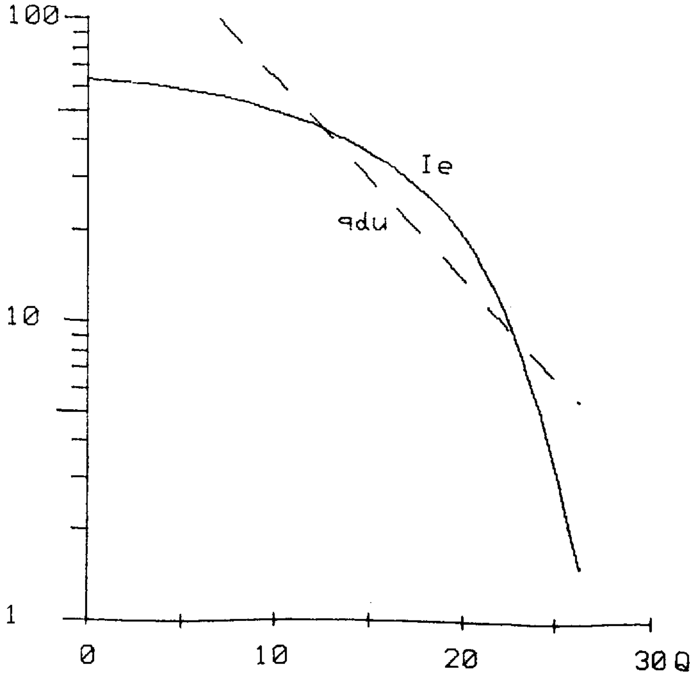


Figure G.3: Equipment impairment factor *Ie* and quantizing distortion qdu as functions of the MNRU quantizing distortion Q

G.7 Equipment impairment factors *Ie* to be used in transmission planning

The databases for deriving the *Ie*-values are rather limited for some of the codecs investigated here. Also, the subjective impression of the low bit-rate codec impairments may vary considerably between subjects. Therefore, some (small) safety margins have been added to the *K*-values in equ. (G.11) as presented in table G.8.

Table G.8: Equipment impairment factors *Ie* = *K* for low bit-rate codecs in transmission planning (Provisional)

Codec	kbit/s	*K*
ADPCM (ITU-T Recommendations G.726 [121] and G.727 [66])	40	2
	32	7
	24	25
	16	50
LD-CELP (ITU-T Recommendation G.728 [81])	16	7
	12,8	20
VSELP (IS 54 [69])	8	20
RPE-LTP (GSM)	13	20
CELP+	6,8	25

G.8 Conclusions

For *transmission planning purposes*, the influence of low bit-rate codecs on speech transmission quality can be evaluated by using *transmission impairment factors*. Each type of codec is associated with a specific *equipment impairment factor Ie*. The influence of several low bit-rate codecs in tandem is characterized by the *sum* of the individual impairment factors for the codecs in the chain.

Annex H: Transmission Quality of Digital Circuit Multiplication Equipment (DCME)

H.1 Introduction

The following is an extract from "Transmission Quality of Digital Circuit Multiplication & Low Rate Encoding", British Telecommunications Engineering, Vol. 12, April 1993 [75], with minor modification.

H.2 Factors affecting transmission quality

The operation of a DCME can be divided into two distinct processes, namely interpolation and low rate encoding. Hence the factors which affect the transmission quality are logically divided into two classes, namely those which are the result of interpolation processes, and those which are the result of the encoding processes. The two classes maybe further subdivided into speech impairments and voiceband data impairments. Only speech impairments are discussed in this text.

H.2.1 Interpolation process impairments

Two possible types of impairment can occur in the interpolation process:

- competitive clipping, which results when the active trunk channel capacity exceeds the available bearer channel capacity; and

- impairments related to activity detection.

Competitive clipping, or freezeout, can easily be reduced to imperceptible levels by ensuring that the loading of the DCME is correct, in terms of the number of channels for a given set of channel occupancy characteristics. This is discussed, at length, in supplements to ITU-T Recommendation G.763 [82] and G.766 [124], and is not discussed further here.

Impairments associated with activity detection result from a number of different causes:

a) noise, echo, crosstalk and extremes of level may result in misoperation, so that speech is clipped, or a silent background is interrupted by bursts of noise or interference;

b) to compensate for the finite detection time of the activity detector, buffering is inserted in the main signal path. This has the benefit that competitive clipping and recognition failures may also be slightly reduced, but the disadvantage that noise contrast effects may be increased. A considerable disadvantage is the increase in the propagation time, hence making echo control a definite requirement;

c) quantisation distortion is increased at the start of the activity due to the fact that the low rate encoding decoder (and possible the encoder as well) has to be reinitialised for each new assignment. This is because after the decoder (and encoder) is connected to the channel it takes some time to collect sufficient samples to model the signal characteristics accurately, typically about 200 ms. This also means that for speech, synchronous tandeming is only likely to be achieved when the DCME is lightly loaded.

Interpolation process impairments for speech:

A known characteristic of speech is that it has a high peak-to-mean ratio (although this is affected by the characteristics of the transducers used in telephones, among other things). Furthermore, speech consists of a mixture of voiced sounds, which have strong tonal components, and unvoiced sounds, which are noise-like, and generally quieter. It also has a syllabic structure, within which individual syllables have a duration of about 300 - 800 ms, with intersyllabic pauses of about 30 - 100 ms.

Since most detection methods are unreliable when working on less than about 10 ms of speech, it is usual to provide 15 - 30 ms of buffering for the main path signal, while detection takes place. This means that initial low energy sounds, which themselves would not register as activity, can still be transmitted through the DCME, provided that a louder sound follows within the buffering period. This is shown in figure H.1.

For similar reasons, a form of hysteresis, known as hangover, is provided, to ensure that the quiet sounds towards the end of an utterance are not suddenly clipped, which would make the speech sound unnatural. When the activity detector first declares the channel inactive the hangover time timer is started. The channel assignment is maintained until the timer has expired. If a new activity is detected within the hangover period, the timer will be restarted from the cessation of that activity. This enables intersyllabic pauses to be bridged, so that speech sounds more natural, with less interruption to breathing noises and room noise. For speech, the hangover time is usually 30 - 100 ms, depending upon the equipment type. Even after the hangover has elapsed, the channel will not necessarily be reallocated immediately. For that to happen there must be sufficient demand, in the form of newly active unallocated channels. If the equipment is lightly loaded, a channel allocation may persist for several minutes after activity ceases.

The noise contrast effect, mentioned under (b) above, is a change in the character of the background noise on the channel when allocation and deallocation occur. It is important that if the real background noise level on the channel is high, then the listening party should receive a similar noise level, irrespective of whether the channel is allocated or not. Furthermore, the noise should ideally be of similar noise level, irrespective of whether the channel is allocated or not. Furthermore, the noise should ideally be of similar character (spectrum, granularity etc.) to the real noise, so that a contrast is not perceived. In some cases, there appears to be benefit in reproducing a lower noise level than the real one, particularly when the level of the real noise is high enough to be annoying, or its character cannot easily be matched (see reference [53]).

H.2.2 Encoding process impairments

The impairments which can occur in the encoding process may be summarised as follows:

- increased quantisation distortion: with advanced encoding methods such as ADPCM, the quantisation distortion tends to be dependant on the predictability of the signal, and therefore may vary as the signal changes;

- variable bit-rate operation: this results in increased quantisation distortion for speech;

- bandwidth restriction: this is a feature of one ADPCM algorithm designed specifically for voiceband data transmission;

- level dependency effects.

Encoding process impairments for speech:

For speech it is difficult to derive an adequate, comprehensive model for the way the ear experiences many of the effects described above. Experience has shown that the most significant effects in most circumstances are likely to be variable bit-rate operation, and level dependency.

Subjective testing has indicated that for the ITU-T Recommendation G.726 [121] algorithms, when the transmitted bit-rate is reduced, by progressively increasing the proportion of samples encoded with 3 bits, rather than 4, the degradation becomes noticeable at an average of about 3,7 bits per sample.

It has also been observed during subjective testing that ADPCM codecs perform best when the speech input level is within a certain range, and when listening level is also within a certain range. These effects depend to some extent both upon the speaker (whether male or female, for example), and upon some non-linearities of the hearing mechanism, and are therefore not easily testable, other than through formal subjective testing.

H.2.3 Transmission quality assessment

When assessing equipment and techniques which are themselves of high complexity, and which moreover are deliberately interactive with the transmissions which are carried through them, it is very important to adopt a structured approach to testing. This means that long before tests are conducted with ‘live’ traffic in an environment which is something like the real network, a judgement must be made about which characteristics are the most critical and therefore deserving of detailed individual examination. It is then down to the skill of the test team to devise meaningful experiments which fully evaluate each characteristic within the range of likely operation. This is a very time-consuming and expensive process, but experience has shown that a more wholesale approach runs the risk of producing results which do not withstand critical scrutiny.

ITU-T Recommendation P.84 [87] gives the subjective testing methodolgy for DCME systems.

H.3 Effects of interpolation process

Figure H.1 shows what happens to speech that is subjected to interpolation. The top trace represents the sequence of events on a particular trunk channel input, the left-hand side of the figure being earliest time. The vertical bars show the short-term average power level (arbitrary scaling). The line labelled Th is the threshold above which the activity detector registers that speech is present. The thick irregular line at the bottom of the trace represents the background noise.

The second trace shows when a bearer channel is allocated to the trunk channel and, by means of different types of shading, for what reason.

The third trace shows the reconstituted trunk channel output, incorporating some artefacts of the interpolation, in the form of noise contrast (exaggerated for clarity) prior to the period A and between H and I, and I and J.

As the DCME loading is increased, less bearer channel time is allocated to the trunk channel. This occurs by the progressive erosion first of the periods D, H and M (when there is really no need for the bearer channel to be allocated to the trunk channel) and then of periods A, E & J (which are provided to give the activity detector sufficient time to operate correctly). At this stage, depending upon the quality of the activity detector, critical listeners may notice some clipping. Finally, when the DCME is operating in severe overload, periods B, F and K, and their associated hangover periods C, G, and L will be affected.

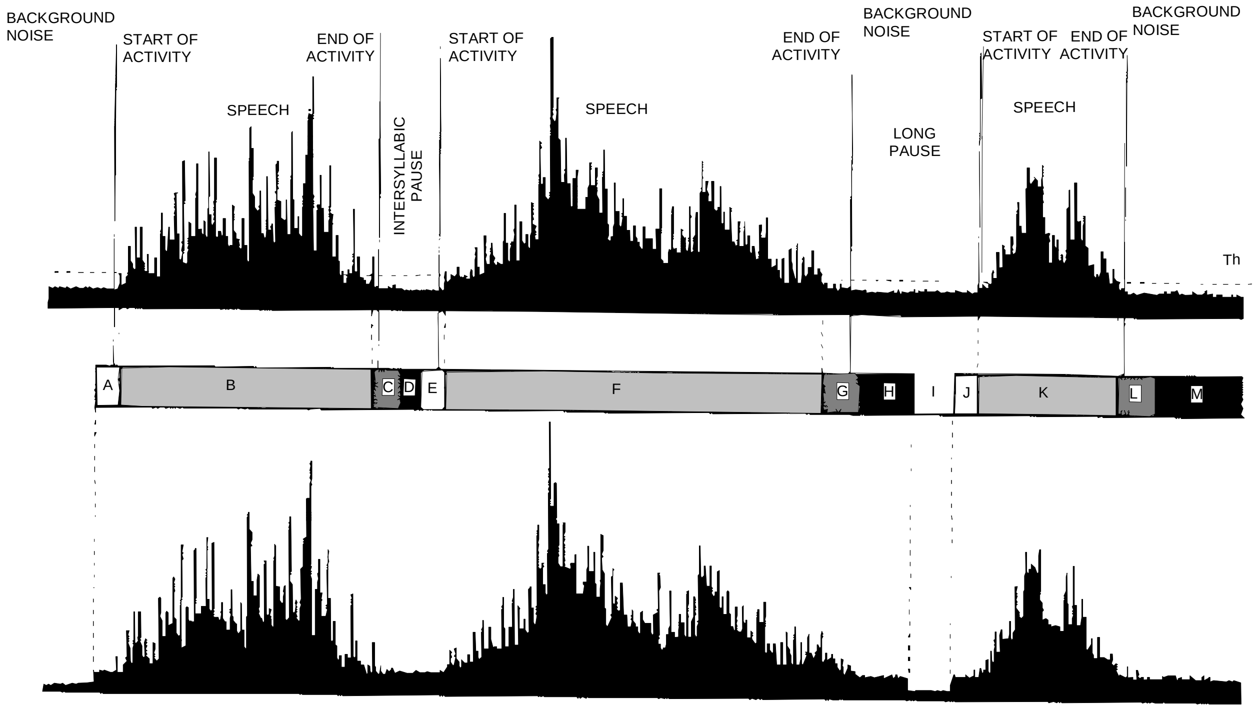


Figure H.1: Explanation of interpolation

Annex J: Usability of telecommunication services; price and quality considerations

J.1 Some definitions

The "usability" of telecommunication services is under study in ISO and ETSI. The fundamental concepts are discussed in reference [96] and some useful definitions are given, part of which are reproduced here.

J.1.1 ISO Draft document CD 9241-11 "Guidance on usability specification and measures" (2 May 1992)

The following definitions are found in ISO draft document CD 9241-11, "Guidance on usability specification and measures" (2 May 1992) [150]:

Usability: The effectiveness, efficiency and satisfaction with which specific users achieve specified goals in particular environments;

Effectiveness: The accuracy and completeness with which specified users can achieve specified goals in particular environments;

Efficiency: The resources expended in relation to the accuracy and completeness of goals achieved;

Satisfaction: The comfort and acceptability of the work system to its users and other people affected by its use.

J.1.2 ETSI enhancements of the definitions

In the present standardisation work, ETSI TC HF and STC HF3 have as far as possible made their definitions consistent with the forthcoming ISO standard. However, ETSI enhances the definitions in the following way:

1) *Usability* is considered a pure ergonometric concept not depending on costs of providing the system. Instead, usability and financial costs together form the concept of *Utility.* This means that an ergonometrically highly usable system may have low utility for a particular user who considers the cost to be too high in relation to his need for using the system.

One advantage of making a distinction between usability and utility is that the actual user (sometimes called the "end user") may not be concerned about costs if he is employed by a company that provides the system for their office work, but he may still be sensitive to the ergonometric usability which affects his work situation and his performance. On the other hand the company, as paying customer of the network operator, has to consider the utility by some kind of cost/benefit analysis, where usability is on the benefit side.

2) *Measures of usability* are assumed to be of two kinds:

a) Performance measures, which are "objective" measures or observations of user behaviour and are focused on task performance, i.e. how well the users can achieve a specific task. Performance measures include both effectiveness and efficiency.

b) Attitude measures, which are "subjective" measures of the users' opinion of working with the system, i.e. how much they like to use the system. Attitude measures include terms such as satisfaction, acceptability, comfort, etc.

J.2 Price and quality considerations

NOTE: "Quality" here means the voice transmission quality of a system relative to what is usually achieved in the wirebound PSTN network.

In the past, customers have not been offered a choice between price and quality, but with the introduction of competition in the provision of voice services, the following issues will arise:

a) network operators will have to choose the level of quality and price for their services in the presence of competition from other network operators and resellers;

b) customers may be able to choose between services with different quality, and will therefore have to make judgements between quality and price.

Because of the absence of competition in the past, there is very little information about customers' perception about this relationship. Information could be obtained by adding questions relating to subjective tests and customer opinion surveys. For example, users could be asked: "Would you be prepared to pay 20 % more for significantly better quality?" or "Would you accept worse quality if the price was reduced by 20 %?"

Information from the market is expected to come first from customers' perception of services offered by resellers where the resale services uses low bit-rate coding to give economies in the use of leased lines.

Annex K: Some notes on the verification of the ETSI computation model

K.1 General

It is not now possible to initiate a major subjective test program to check from scratch all facets of the ETSI model. For economic reasons, it would be impossible to match the extensive subjective investigations on which the Bellcore Transmission Rating Model or the BT CATNAP model are based. However, it is hardly necessary to "verify" the ETSI model by such ambitious efforts.

Some aspects of transmission impairments are so well known that their inclusion in the ETSI model should not need verifying by new subjective tests. Examples of such transmission parameters are:

- overall Loudness Rating and circuit noise;
- qdu for normal PCM;
- talker sidetone as such;
- talker echo for delays longer than 10 ms. (Note that "limit curves" in CCITT Recommendation G.131 [14] have proven to give adequate talker echo performance in practical applications).

Other transmission parameters would benefit from extended subjective tests in order to increase the confidence in the use of the model. Examples are:

- talker echoes for delays shorter than 10 ms;
- impressions of speech quality of low bit-rate codecs for longer conversations than just 10 seconds speech samples, and also with burst errors, distributed over a longer time;
- the various effects associated with mobile communications;
- combination effects of impairments, in particular when new impairments are involved.

As discussed shortly in what follows, the "verification" process should ideally include several procedures, namely formal subjective tests under controlled condition, analysis of customer surveys with regard to speech communication quality and collection of experiences from application of the model in transmission planning.

The process should:

1) identify suitable subjective test databases for verification;
2) supplement the existing subjective test results with new tests, in particular for combinations of impairments;
3) identify test criteria, e.g. error limits;
4) test and compare the model with the subjective test database;
5) identify the model's strengths and weaknesses with recommendations for future improvements/enhancements.

One must of course be aware that subjective tests and surveys have statistical variability so that the computation results from the model cannot be expected to match exactly each and every subjective result. (In general, differences in MOS, Mean Opinion Scores, may amount to 0,2 or 0,3 MOS units between tests repeated under similar conditions with a representative number of subjects. However, sometimes differences of up to 0,5 MOS can be found). Even if it is not expected that the ETSI model will give an exact match with subjective test results of other recognized transmission planning models, it should however display the same general characteristics. It should also be within the normal variance expected when comparing results from different subjective tests. (It is recommended that both graphical and numerical comparisons are used, enabling the extent of any differences to be quantified).

Formal subjective tests may rather easily be designed to produce results which have a high *reliability*, but a high *validity* can be somewhat more difficult (although not impossible) to achieve as considerable care in the test design is needed. For results from customer surveys, the opposite is true.

Note that Annex E contains an overview of subjective test methods.

K.2 Formal subjective tests

Formal subjective tests should be conducted with the usual appropriate care. (Note that annex E is only a tutorial overview of subjective tests, not a handbook of how to conduct subjective tests). Conversational tests are in general to be preferred to listening-only tests unless, for certain cases, it is known from experience that they do not differ significantly. It is very desirable to include one or more normalised "reference impairment conditions" in tests.

K.3 Analysis of customer surveys

In general, the analysis of customer surveys requires much more effort. A big problem is to associate the interviewed customer with the parameters of the connection involved. A certain guidance can be had from the investigations made by AT&T (SYBIL, etc.). Also, theoretically, the output from an INMD (In-service, Non-intrusive Measuring Device) could be employed to a certain extent.

K.4 Experiences from the use of the model in transmission planning

A vast experience exists of transmission planning and the feed-back of users' reactions. It should be fruitful to test the model on various real circuits and connections and see to what extent "critical cases" as well as "acceptable cases" can be identified.

Annex L: Propagation delay and echo control procedures

The propagation delay in telecommunication networks affect both computer communication protocols and audiovisual communication between humans. The most noted effect is the echo problem for telephony which depends heavily on the propagation delay of the connection.

Note that the propagation delay determination signalling procedure and the echo control signalling procedures are in principle separate procedures in the network. However, the output from the propagation delay determination procedure can in some cases trigger echo control procedures.

L.1 Propagation delay determination in the switching network

L.1.1 Availability of propagation delay determination procedures

Access protocol (e.g. DSS1) signalling:

There is no propagation delay determination signalling procedure in any of the access protocols.

R2, #5, TUP, ISUPv1 (ITU-T Recommendation Q.767 [149]) signalling:

There is no propagation delay determination signalling procedure.

ISUPv2 (ITU-T Recommendation Q.764 1993 [148]) signalling:

There is specified a propagation delay determination signalling procedure. However, most network operators do not implement the ISUP 1993 propagation delay determination procedure.

BISUP (ITU-T Recommendation Q.2764 [46]) signalling:

There is a propagation delay determination signalling procedure.

L.1.2 Propagation delay detemination signalling protocol description

The propagation delay determination procedure in ETSI ISUP v2 calculates the accumulated propagation delay for each call. The resolution is 1 ms.

The exchanges receive an accumulated propagation delay value from the previous exchange, add the propagation delay for the route to the succeeding exchange, and send the new accumulated delay value to the succeeding exchange.

Input to the procedure is assumed propagation delay indicated by data in the exchange. The data shall indicate approximate propagation delay on the route to the succeeding exchange. The data is set per route, and the resolution is 1 ms. Data for propagation delay can be calculated and loaded manually or by the Telecommunication Management Network (TMN), depending on the implementation.

Propagation delay per route can be considered as static data, and not calculated per call.

Proposal:

For the switching network, an important quality of a circuit is the propagation delay. This means that propagation delay should preferably be defined as an attribute of the traffic path classes of the information model for the TMN application for the switched network.

L.2 Need for a propagation delay determination procedure in the transmission (transport) network

The switching network, consisting of exchanges and circuits (traffic paths), can be considered to be a client of the transmission or transport network, consisting of line transmission systems, cross connects and multiplexers.

The products delivered by the transport network to the switching network are transmission paths, which are seen as circuits by the switching networks. Propagation delay can be considered as one of the important qualities of this product.

Today there are no standardised procedures for handling propagation delay aspects in the transmission (transport) network. This is not a satisfying situation.

Proposal:

The propagation delay should be defined as an attribute to the transmission path object classes in the information (semantic) model of the transport network.

A TMN application should map the calculated propagation delay of the transmission path to the corresponding traffic path of the TMN application for the switching network.

The propagation delay value of the transmission path will depend upon both the values of propagation delay in the line sections, and the configuration of the transmission network.

Consequently, the TMN application for the transport network should register the propagation delay for each line section, and calculate the accumulated end-to-end propagation delay for each transmission path for each multiplexing level, taking into account the current configuration of the transmission (transport) network.

The propagation delay for the line systems would be static data, calculated/measured and loaded manually. The propagation delay for transmission paths on other multiplexing levels or cross connect levels must be recalculated by the TMN application for each reconfiguration of the transmission network. The resolution should be 0,1 ms in order to get sufficient accuracy in the accumulated value.

L.3 Echo control procedure

L.3.1 Echo control requirements

The general requirements for end-to-end echo attenuation are described in CCITT Recommendation G.131 [14].

ITU-T Recommendation Q.115 [42] describes how the requirements of ITU-T Recommendation G.131 [14] shall be applied in the different networks.

Echo control signalling procedures for each protocol are described in the relevant recommendations for the protocols.

L.3.2 Bearer service dependency

Echo attenuation is an important quality of service for the users of telephony calls.

Two types of telephony are defined for public telecommunication networks, 3,1 kHz telephony and 7 kHz telephony.

For 7 kHz telephony, echo attenuation is provided by the user terminal equipment.

The 3,1 kHz telephony can be provided by user terminals in two different ways:

- the terminal equipment can ask the network to provide a transparent computer communication connection, e.g. 64 kbit/s unrestricted bearer service, and use this connection to provide the telephony teleservice. In this case the user terminal must also provide the necessary echo attenuation;

- the terminal equipment can ask the network to provide the speech or the 3,1 kHz audio bearer service. This is the most common case, and e.g. for calls terminating in analog subscriber lines the bearer service is always 3,1 kHz audio. In this case the network must provide the necessary echo attenuation;

- ISDN and B-ISDN networks can also provide the multiuse bearer service. In this case the network must provide the necessary echo attenuation when fallback to speech/3,1 kHz audio information transfer capability has been initiated.

The need for echo attenuation provided by the network depends upon the requested bearer service, but not which type of telecommunication network is used to provide it, e.g. PSTN, ISDN or a different computer communication network.

L.3.3 End-to-end transmission quality requirements

The general requirements for echo attenuation are described in CCITT Recommendation G.131 [14]. The need for echo attenuation is dependent on the values of a number of parameters. The most important parameter is the propagation delay in the transmission path of the connection.

- for low values of propagation delay, there is no critical requirement for echo attenuation, except for loudspeaking telephones. Additional echo attenuation needs not to be provided by the network;

- for medium values of propagation delay, the need for additional echo attenuation is depending upon the actual values of a number of parameters;

- for large values of propagation delay, there is always a need to provide additional echo attenuation.

CCITT Recommendation G.131 [14] gives a basis for deciding echo attenuation requirements.

L.3.4 Determination of need for echo cancellers

The values of the different CCITT Recommendation G.131 [14] parameters are usually not known by user terminal equipment. For user terminal equipment there are usually two relevant options:

a) always provide echo cancellers, or

b) request a bearer service for which the network provides echo cancellers when necessary.

The network usually does not know the actual values for most CCITT Recommendation G.131 [14] parameters. However, the network usually has some knowledge of the dominant parameter, the propagation delay.

Because the network does not know the values of CCITT Recommendation G.131 [14] parameters for an actual call, the requirements of CCITT Recommendation G.131 [14] cannot be directly applied in the network. Instead some implementable echo control requirements must be specified for different networks by a reasonable application of CCITT Recommendation G.131 [14] requirements. ITU-T Recommendation Q.115 [42] describes requirements for inserting echo control devices in PSTN, ISDN and B-ISDN networks.

The requirements of ITU-T Recommendation Q.115 [42] are based on assumptions about network configuration and call routing statistics. They are also depending upon the limitations of the available signalling protocols.

According to CCITT Recommendation G.131 [14] the need for echo control devices can be calculated separately for each transmission direction. Requirements for echo control in ITU-T Recommendation Q.115 [42] are always symmetric. Either no echo control devices are required, or both an incoming and an outgoing echo control device are required.

L.3.5 Location of echo control devices

Location of echo control devices are decided based on economic considerations and the requirements of CCITT Recommendation G.131 [14] and ITU-T Recommendation Q.115 [42].

In most PSTN and ISDN networks echo control devices are not located in local exchanges. Outgoing echo control devices are usually located in outgoing international gateway exchanges. Incoming echo control devices are usually located in incoming international gateway exchanges. Incoming echo cancellers are located in outgoing digital mobile telephone gateway exchanges. Outgoing echo cancellers are located in incoming digital mobile telephone gateway exchanges. Echo cancellers are located in GSM mobile handsets. In some large countries, echo control devices are also located in national transit exchanges.

Because of the additional delay in B-ISDN networks, it is expected that also some local B-ISDN exchanges will have echo cancellers. Emulation of the 64 kbit/s connection in ATM cells causes an additional delay approximately equivalent to propagation delay in a 1 200 km fibre optic cable.

L.3.6 Echo control signalling protocol aspects

Access protocols (e.g. DSS1) signalling:

There is no echo control signalling procedure in any of the access protocols.

#5 signalling:

No echo control signalling procedure is available. The exchanges having echo control devices mainly make decisions concerning echo control or not based on data associated with the route used for the call, e.g. a satellite route or another "long" route. There are no considerations for supplementary services, e.g. Call transfer, Conference call, Call diversion etc.

R2, TUP, ISUPv1 (ITU-T Recommendation Q.767 [149]) signalling:

Echo control signalling procedures with limited functionality are available. The exchanges having echo control devices mainly make decisions concerning echo control or not based on data associated with the route used for the call, e.g. a satellite route or another "long" route. There are no considerations for supplementary services, e.g. Call transfer, Conference call, Call diversion etc.

ISUPv2 (ITU-T Recommendation Q.764 1993 [148]) signalling:

There is a "dynamic" echo control signalling procedure. The exchanges can make decisions concerning echo control based on the output from the propagation delay determination signalling procedure, as well as from data per route or digit analysis. When the accumulated propagation delay exceeds a predefined value, e.g. 25 ms, the exchange should initiate the echo control signalling procedures.

That the procedure is "dynamic" means that there is a negotiation process in both directions between the exchanges. The need for echo control and possible availability of echo control devices are indicated in the signalling. Echo cancellers are inserted in the exchanges having echo cancellers available and being located closest to the ends of the connection. The exchanges can make considerations for supplementary services.

However, most network operators do not implement the ISUP 1993 echo control signalling procedure.

BISUP (ITU-T Recommendation Q.2764 [46]) signalling:

An echo control signalling procedure has been proposed, but not yet approved in ITU-T SG11. With this procedure, the exchanges can make decisions concerning echo control based on the output from the propagation delay determination procedure. When the accumulated propagation delay exceeds a predefined value, e.g. 25 ms, the exchange should initiate the echo control signalling procedures. The proposed procedure is "dynamic". The exchanges can make considerations for supplementary services.

Annex M: Aspects of voice transmission quality which need further studies

In the preparation of this ETR, some aspects of voice transmission quality were put aside due to lack of results from subjective studies in the area. This annex lists a number of these aspects which have not been included in the model, but might be taken into account if it is found that these aspects have a definitive impact on the speech quality, as perceived by the user. Further studies of these aspects are therefore called for.

M.1 Treatment of multiple echoes

M.1.1 Background

In mixed analogue/digital networks with 2-wire interfaces between the different parts of a connection or between public and private networks, more than one 4-wire loop may exist. All equipment converting from 4-wire to 2-wire (usually hybrids) are basically a source for return signals. In those configurations, several echo paths with different echo path delays may cause multiple echoes to the talker. Existing planning rules for talker echo are usually based on a delay/loss ratio, derived from investigations and subjective tests on only one echo path. The planning can be performed using the formula for talker echo loudness rating TELR, as given in subclause 6.5.6.2 of this ETR.

The configurations to be considered will mainly consist of one 4-wire loop with a low value of echo path delay but with low loss, i.e. low values for TELR, and a second echo path with higher delay and usually also higher values for the TELR. All echo paths will be in the range of 25 ms for the mean one-way transmission time, so that no echo control devices are used. The usual planning method is to perform a separate investigation for each echo path, comparing the expected TELR with the required TELR, according to the formula in subclause 6.5.6.2 of this ETR. If one of the considered echo paths provides a TELR well above the calculated TELR, this path can be neglected for further planning.

The problem arises when all, or only two echo paths provide each a TELR close to the required TELR. In these cases an additional unknown degradation of speech quality must be expected. In contrary, a 4-wire loop close to the talker with a very short delay, or the sidetone of the talker's telephone set, could provide an effect of masking to the remaining echo of all other paths. Previous investigations to this effect should be continued more in detail.

M.1.2 Required investigation

The investigation should preferably result in a modified calculation of the respective impairment factor in the ETSI-model, and/or in a modified formula in subclause 6.5.6.2 of this ETR. The planning rules with respect to echo must be extended with a guidance for those situations. A limit should be defined as a difference between actual TELR and required TELR, to identify those echo paths which can be neglected during echo planning. These investigations mainly based on calculations should have, as to our opinion, a confirmation by subjective tests.

The masking effect to network echo by sidetone or 4-wire loops with short delay can only be done by subjective tests. The investigations should be based on a test arrangement providing an echo path with a one-way transmission time in the range of 10 ms to 25 ms, and a STMR value on the range of 10 dB to 20 dB. If these subjective tests confirm the effect of masking in a sufficient order, the results should be integrated in the respective impairment factors of the ETSI-model.

M.2 Weighting algorithm for acoustic echo paths

M.2.1 Background

For the consideration of echo effects and the respective planning, the echo path loss, mainly of the terminating hybrids, is derived from the integral of the power transfer characteristics A(f), weighted by a negative slope of 3 dB per octave in the range from 300 Hz to 3 400 Hz (see also subclause 6.5.2.3 of this ETR). The respective methods for calculation (trapezoidal rule), given in ITU-T Recommendation G.122 [25], annex B, section 4, have been common practice for several years.

However, in principle this method is only valid for echo path losses with a frequency shape as usually exist in pure electrical paths. The increasing use of digital equipment in public and private networks will provide in future more and more fully 4-wire routed connections, where the acoustic path of the telephone sets between receiver and transmitter capsule, or between loudspeaker and microphone in case of handsfree mode, will act as the only source for return signals. In contrary to pure electrical paths, the frequency response characteristics of these acoustic paths differ in a high amount.

The characteristic of this acoustic path of a telephone set with respect to echo is expressed with the term Terminal Coupling Loss (TCL), and presently weighted with the same algorithm as for pure electrical paths. It seems necessary to confirm this practice, or to develop - if necessary - a modified weighting algorithm.

M.2.2 Required investigation

The confirmation, if the weighting method as given in ITU-T Recommendation G.122 [25], annex B, section 4 is also applicable for acoustic echo paths, should be obtained by subjective comparison between a pure electrical echo path with known weighted echo loss and a digital telephone set with an equal value for the TCLw. The judgement should be referred only to echo effects. The investigation must include acoustic paths in the handset mode and in the handsfree mode, where the values for echo loss or TCLw during test should be chosen accordingly. The frequency response of capsules or microphones and loudspeakers and the SLR/RLR-values of the telephone sets in the test arrangement should be within specified limits.

M.3 Echo cancellers in conjunction with non-linear echo paths

M.3.1 Background

To avoid the disturbing effect of talker echo, mainly in international connections with high values of one-way transmission time, echo cancellers (ECs) within the public networks are used. Usually these ECs are located in the international switching centers (ISCs), providing the access to international traffic. The remaining national part of the connection between the ISC and the first point with a conversion from 4-wire to 2-wire (hybrid in a local exchange) forms the echo path of the EC. Usually all those ECs are designed to meet the specifications on ITU-T Recommendation G.165 [19].

If private networks are connected digitally with the public network, the echo path is extended by the routing within the private network, terminated with the acoustic path of a digital telephone set or a possible 4-wire to 2-wire conversion in case of analogue telephone sets. The links between the exchanges of the private networks are mainly digital leased lines. Due to economical reasons, these links are more and more equipped with speech companding devices using ADPCM or other methods to increase the number of channels. The echo path for the EC in the ISC will provide not only higher values for the delay, but has now become non-linear, in contrary to the assumption in ITU-T Recommendation G.165 [19]. Therefore a degradation of the operational characteristics of the EC must be expected.

M.3.2 Required investigation

The investigation must consider two different facts: the insertion of non linearity, and an increase in the echo path delay. It is strongly recommended to perform these investigations by subjective tests instead of pure electrical measurements using white noise as in ITU-T Recommendation G.165 [19]. The test arrangement must contain ECs in accordance with ITU-T Recommendation G.165 [19], including non linear processing (NLP) and the simulation of an echo path with inserted speech companding devices and a one-way echo path delay in the range of 25 ms to 30 ms. The echo path in the test arrangement is terminated by a hybrid providing an echo loss of 10 dB.

It seems necessary to perform these subjective tests for different types of speech companding devices as they are presently in use. The results should preferable be available as a MOS or percentage POW rating. Depending on the results (only degradation of the cancelling characteristics or fully inoperative) also a modification of the respective impairment factor in the ETSI-model is requested.

The investigation should also identify if an encountered degradation of the EC is caused by the non-linearity or the delay of the echo path.

M.4 Echo cancellers in conjunction with acoustic echo paths

M.4.1 Background

The background for this problem is similar to the case of echo cancellers (ECs) in conjunction with nonlinear echo path (subclause M.3). In this case, a pure electrical echo path is replaced by the acoustic path of a digital telephone set in conjunction with higher values for echo path delay. ECs, designed according to ITU-T Recommendation G.165 [19] assume pure electrical echo paths with the respective frequency responses; therefore a degradation of the operational characteristics can be expected also in those configurations.

M.4.2 Required investigation

In contrary to the test arrangement in subclause M.3, the simulation of the echo path must only provide a termination by a digital telephone set and a one-way delay in the range of 25 ms to 30 ms, but no speech companding device. The tests should as well be performed on a subjective basis, expressed in MOS or percentage POW rating, including the identification if an encountered degradation is caused by the characteristics of the acoustic echo path, or by an exceeding of the EC's delay range.

The tests must be performed separately for the digital telephone set in the handset mode and in handsfree mode.

M.5 Mean one-way transmission time exceeding 400 ms

M.5.1 Background

International connections will be routed more and more via satellite links with one-way transmission times of 260 ms. Additionally, the use of DCME, low bit-rate codes, processing times, etc., within public networks, will result in transmission times close to, or already exceeding, the limit of 400 ms, as stated in ITU-T Recommendation G.114 [36]. In the field of business communications, those connections are terminated by private networks, using the same modern technology and as well subject to economical solutions. This will lead to generally higher values for the mean one-way transmission time and an increase of those connections where the limit of 400 ms is exceeded. Therefore difficulties with respect to conversation dynamics also on echo-free connections must be expected.

Respective investigations and subjective tests, performed mainly in USA, as they can be found as a summary in annex B to ITU-T Recommendation G.114 [36] lead partly to different results and can be used only for specific applications. This summary however is important with respect to the issue that "the effects of pure delay (no echo) on conversation dynamics can be detected well below 400 ms one-way delay if ... highly interactive tasks are used".

As to our opinion, it seems necessary to deviate in our ETR from the issue of a single absolute upper limit, for the benefit of the network operator. It should be possible and in the responsibility of the operator to use different upper limits depending on specific applications (e.g. normal speech transmission or parallel sound to video transmission), the structure of the network, its interconnection with other networks and mainly the constraint for economical solutions. Of course, sufficient information and guidelines about the expected difficulties on a conversation must be made available in the ETR, to support the operator's decision.

M.5.2 Required investigation

The decision, how to perform investigations and subjective tests on echo free connections with respect to transmission times exceeding 400 ms is very difficult and needs further discussion. One of the questions is what upper limit for the delay should be included. The major problem however is, how to perform the subjective tests: listener tests only or conversation tests using different tasks, and should the judgement be done only by a MOS rating or separating between interruptability and further factors influencing interactive tasks.

As already stated in the background, it seems no more longer possible to issue a single absolute upper limit for the mean one-way transmission time. The investigation should therefore contain the evolution of guidelines for the upper limit, depending on different factors. For planning purposes, the given limits should be expressed all in the same unit MOS or percentage POW, independent of their relation to a specific factor, describing the influence to conversation quality like interruptability, etc.

It is not intended to try an integration of the results into the ETSI-model. As to our opinion, the difficulties for conversation dynamic arising with high values of one-way transmission time, can be considered and judged separately from all other transmission parameters.

M.6 Asymmetry of loss

M.6.1 Background

In ITU-T Recommendation G.121 [38], clause 2.2 it is noted that "it has been found practical to introduce a certain difference in loss between the directions 4-wire-to-2-wire and 2-wire-to-4-wire". This difference is obtained through the so-called "R- and T-pads", and is usually in the order of 6 - 7 dB. It is assumed that administrations (countries) adopt this principle in their national networks. In an international connection, the loss will then be of equal value in both transmission directions.

For a number of reasons there may be a need to insert extra loss (or gain) in one transmission direction. However, according to ITU-T Recommendation G.121 [38], "the difference in loss between the two directions of transmission on an international connection should not exceed 8 dB, preferably not 6 dB". As a result of the liberalisation of telephony services, a call may be passing through several different networks, each having its own transmission planning. In such a scenario, it may be difficult to ensure that the loss difference is kept within the "ITU-T limits".

M.6.2 Required investigation

Some investigations have already been made, see reference [97]. It has been found that the user's behaviour emphasizes the loss asymmetries. There is a tendency for users to speak more loudly when the signal is weak, and more quietly when the signal is too loud.

At present, the influence of OLR (too high - too low) is included in the ETSI model, but not any possible quality degradation perceived by users due to loss asymmetry. This would require further studies.

M.7 Sidetone

M.7.1 Background

In the past sidetone measurements were performed only at analogue telephones because digital sets were not available. Furthermore tests with digital terminals need a network simulation for the test configuration which requires more expenditure on equipment. The last argument is not a reason for not performing the tests but it does not make the testing any easier and has lead recently to the situation that the values for digital sets are derived from the experiences of analogue sets by consideration of the special features of digital sets. But the tests for digital sets should be performed now for checking and updating the recommended values because the electrical sidetone paths are different betweeen both terminals. That concerns particularly the frequency response; for analogue sets this characteristic is defined mainly by the hybrid with the used termination whereby level and frequency response are coupled and not adjustable independently. But for digital sets the possibility for independent adjustment of level and frequency response for the sidetone path is given principally and furthermore the sidetone path is independent from the sending and receiving circuits. That allows the implementation of calculated lower sidetones which otherwise, may be more influenced by the mechanical handset characteristic (coupling, shape).

M.7.2 Required investigations for Talker Sidetone

- earlier subjective tests in 1980 made by AT&T vary the loudness and the noise, echo path loss and echo path delay, frequency response above and below 1 kHz. The results have been used to derive an extension to the AT&T opinion model. But the sidetone is measured at analogue sets in EARS and not as STMR which should be repeated and tested additionally with digital sets;

- the preferred range and the accepted range as it is stated in CCITT Supplement No.11 [41] is based on talking-only tests and should be repeated, and needs to be confirmed by conversational tests. Furthermore the influence of noise (room noise, circuit noise) should be studied also and the variation of the speech level should be controlled additionally as a function of STMR and room noise;

- the masking effect by more loud sidetone for reflected and delayed speaker signals should be tested more in detail and with better solution for the round trip delay in the range up to 4 ms as it is presented by test results of the Swedish Telecom in 1989.

M.7.3 Required investigations for Listener Sidetone

- subjective tests for satisfactory LSTR according figure 3 in CCITT Supplement No.11 [41] should be repeated but:

    - with analogue and digital telephones;

    - with the direct measurement of room noise sensitivity S_{RNST} of the diffuse sound field and with calculation of LSTR according ITU-T Recommendations P.79 [33] and P.64 [40] and not using the approximation DELSM;

    - with different room noise levels;

    - with conversational test conditions.

- the enhancement of room noise suppression should be tested subjectively as a result of different types of implemented non-linear gain characteristics and compared with the measured LSTR for the decision whether the principle of measurement of LSTR is correct applicable for this implementation or not;

- it should be tested whether the LSTR properly takes into account room noise leaking past the earcap of low acoustic impedance receivers;

- for linear telephone sets further details of the relationship between STMR and LSTR should be tested regarding different shapes of the handset and different characteristics of the microphone revealing the correct use of DELSM / D;

- for common non-linear gain characteristics used in both of the send and the sidetone paths it should be tested whether DELSM is applicable for the LSTR-STMR difference.

Annex N: Bibliography

- B.F. Barry, A.D. Wallis: "New Techniques for the Measurement and Rating of Impulsive Noise", 8th FASE Symposium on Environmental Acoustics, page 291 (1989);

- C.G. Rice: "Annoyance due to Impulse Noise", CEC Studies: Final Report - ISVR Memorandum n., page 690 (1989);

- CCITT Recommendation G.152 (1988; Characteristics approbriate to Long-distance Circuits of a Length not exceeding 2500 km;

- CCITT Recommendation G.171 (1988; Transmission Plan Aspects of privatly operated Networks;

- CCITT Recommendation G.212 (1988; Hypothetical Reference Circuits for analogue Systems;

- CCITT Recommendation G.215 (1988; Hypothetical Reference Circuit of 5000 km for analogue Systems;

- CCITT Recommendation G.221 (1988; Overall Recommendations relating to Carrier-Transmission Systems;

- CCITT Recommendation G.230 (1988; Measuring Methods for Noise produced by Modulating Equipment and through Connection-Filters;

- CCITT Recommendation G.311 (1988; General Characteristics of Systems providing 12 Carrier Telephone Circuits on an Open-wire Pair;

- CCITT Recommendation G.441 (1988; Permissible Circuit Noise on Frequency-division Multiplex Radio-relay Systems;

- CCITT Recommendation G.444 (1988; Allowable Noise Power in the hypothetical Reference Circuit of trans-horizon Radio-relay Systems for Telephony using Frequency-division Multiplex;

- CCITT Recommendation G.445 (1988; Allowable Noise Power in the hypothetical Reference Circuit for Frequency-division Multiplex Telephony in the fixed Satellite Service;

- CCITT Recommendation G.471 (1988; Conditions necessary for Interconnection of Mobile Radiotelephone Stations and International Telephone Lines;

- CCITT Recommendation G.473 (1988; Interconnection of a Maritime Mobile Satellite System with the International automatic switched Telephone Service; Transmission Aspects;

- ITU-T Recommendation P.10, 1993; Vocabulary of Terms on Telephone Transmission Quality and Telephone Sets;

- CCITT Recommendation P.48 (1988; Specification for an Intermediate Reference System;

- CCITT Recommendation P.73, Orange Book V, 1977; Measurements of Sidetone Reference Equivalent;

- CCITT Rev. Recommendation O.71, Volume IV, Blue Book 1988; page 109-112, Impulsive Noise Measuring Equipment for Telephone-type Circuits;

- CCITT Supplement No. 20 to Section I of the Series G Recommendations, Blue Book 1988; Possible Combinations of Basic Transmission Impairments in hypothetical Reference Connections;

- CCITT Supplement No. 21 to Section I of the Series G Recommendations, Blue Book 1988; The Use of quantizing Distortion Units in the planning of International Circuits;

- CCITT Supplement No. 22 to Sections 2 to 5 of the Series G Recommendations (1988; Mathematical Models of Multiplex Signals;
- CCITT Supplement No. 23 to Sections 2 to 5 of the Series G Recommendations (1988; Explanatory Notes for the Information of Designers of a Maritime Mobile Satellite System;

- CCITT Supplement No. 24 to Section I of the Series G Recommendations, Blue Book 1988; Consideration of quantizing Distortion Units of some Digital Devices that process encoded Signals;

- CCITT Supplement No. 27 to Sections 2 to 5 of the Series G Recommendations (1988; Interference from external Sources;

- Supplement 29 to ITU-T Series G Recommendations: "Planning of mixed analogue-digital circuits (chains, connections)", 1993;

- CCITT Supplement No. 4 to Sections 2 to 5 of the Series G Recommendations (1988; Certain Methods of avoiding the Transmission of exessive Noise between interconnected Systems;

- CCITT Supplement No. 6 to Sections 2 to 5 of the Series G Recommendations (1988; Example showing how the total Value of Line Noise specified for hypothetical Reference Circuit on Open-wire Lines might be broken down into its various Components;

- ETSI VTQ M-E, TD 92/6, "A Model for Evaluation of the Combination effects of Impairments", Chairman VTQ M-E, April 1992;

- Handbook on Transmission Planing (Geneva 1987);

- S. Wong "An Evaluation of 6.4 kbit/s Speech Codecs for IMMARSAT-M System", ICASSP ´91, Paper S9.16, pp. 629-632;

- COM XV-D.465: "Evaluation of LD-CELP Performance for Use in DCME", COMSAT, INTELSAT, Nov. 1992;

- Report on ITU-T SG 12 Question 13/12 - Methods for Measuring and Modelling the Effects on Non-linear Processes on the Speech Quality of Transmission Systems, Rapporteurs: N. Kitawaki (Japan), D.J. Atkinson (USA), Feb. 1994.

History

Document history	
July 1996	First Edition

ISBN 2-7437-0768-2
Dépôt légal : Juillet 1996