INTERNATIONAL TELECOMMUNICATION UNION

CCITT

THE INTERNATIONAL
TELEGRAPH AND TELEPHONE
CONSULTATIVE COMMITTEE

BLUE BOOK

VOLUME V

TELEPHONE TRANSMISSION QUALITY

SERIES P RECOMMENDATIONS

IXTH PLENARY ASSEMBLY
MELBOURNE, 14-25 NOVEMBER 1988

Geneva 1989

INTERNATIONAL TELECOMMUNICATION UNION

CCITT

THE INTERNATIONAL
TELEGRAPH AND TELEPHONE
CONSULTATIVE COMMITTEE

BLUE BOOK

VOLUME V

TELEPHONE TRANSMISSION QUALITY

SERIES P RECOMMENDATIONS

IXTH PLENARY ASSEMBLY
MELBOURNE, 14-25 NOVEMBER 1988

Geneva 1989

ISBN 92-61-03441-1

CONTENTS OF THE CCITT BOOK
APPLICABLE AFTER THE NINTH PLENARY ASSEMBLY (1988)

BLUE BOOK

FASCICLE III.6 – Line transmission of non-telephone signals. Transmission of sound-programme and television signals. Series H and J Recommendations (Study Group XV).

FASCICLE III.7 – Integrated Services Digital Network (ISDN) – General structure and service capabilities. Recommendations I.110-I.257 (Study Group XVIII).

FASCICLE III.8 – Integrated Services Digital Network (ISDN) – Overall network aspects and functions, ISDN user-network interfaces. Recommendations I.310-I.470 (Study Group XVIII).

FASCICLE III.9 – Integrated Services Digital Network (ISDN) – Internetwork interfaces and maintenance principles. Recommendations I.500-I.605 (Study Group XVIII).

Volume IV

FASCICLE IV.1 – General maintenance principles: maintenance of international transmission systems and telephone circuits. Recommendations M.10-M.782 (Study Group IV).

FASCICLE IV.2 – Maintenance of international telegraph, phototelegraph and leased circuits. Maintenance of the international public telephone network. Maintenance of maritime satellite and data transmission systems. Recommendations M.800-M.1375 (Study Group IV).

FASCICLE IV.3 – Maintenance of international sound-programme and television transmission circuits. Series N Recommendations (Study Group IV).

FASCICLE IV.4 – Specifications for measuring equipment. Series O Recommendations (Study Group IV).

Volume V – Telephone transmission quality. Series P Recommendations (Study Group XII).

Volume VI

FASCICLE VI.1 – General Recommendations on telephone switching and signalling. Functions and information flows for services in the ISDN. Supplements. Recommendations Q.1-Q.118 *bis* (Study Group XI).

FASCICLE VI.2 – Specifications of Signalling Systems Nos. 4 and 5. Recommendations Q.120-Q.180 (Study Group XI).

FASCICLE VI.3 – Specifications of Signalling System No. 6. Recommendations Q.251-Q.300 (Study Group XI).

FASCICLE VI.4 – Specifications of Signalling Systems R1 and R2. Recommendations Q.310-Q.490 (Study Group XI).

FASCICLE VI.5 – Digital local, transit, combined and international exchanges in integrated digital networks and mixed analogue-digital networks. Supplements. Recommendations Q.500-Q.554 (Study Group XI).

FASCICLE VI.6 – Interworking of signalling systems. Recommendations Q.601-Q.699 (Study Group XI).

FASCICLE VI.7 – Specifications of Signalling System No. 7. Recommendations Q.700-Q.716 (Study Group XI).

FASCICLE VI.8 – Specifications of Signalling System No. 7. Recommendations Q.721-Q.766 (Study Group XI).

FASCICLE VI.9 – Specifications of Signalling System No. 7. Recommendations Q.771-Q.795 (Study Group XI).

FASCICLE VI.10 – Digital subscriber signalling system No. 1 (DSS 1), data link layer. Recommendations Q.920-Q.921 (Study Group XI).

FASCICLE VI.11 – Digital subscriber signalling system No. 1 (DSS 1), network layer, user-network management. Recommendations Q.930-Q.940 (Study Group XI).

FASCICLE VI.12 – Public land mobile network. Interworking with ISDN and PSTN. Recommendations Q.1000-Q.1032 (Study Group XI).

FASCICLE VI.13 – Public land mobile network. Mobile application part and interfaces. Recommendations Q.1051-Q.1063 (Study Group XI).

FASCICLE VI.14 – Interworking with satellite mobile systems. Recommendations Q.1100-Q.1152 (Study Group XI).

Volume VII

FASCICLE VII.1 – Telegraph transmission. Series R Recommendations. Telegraph services terminal equipment. Series S Recommendations (Study Group IX).

FASCICLE VII.2 – Telegraph switching. Series U Recommendations (Study Group IX).

FASCICLE VII.3 – Terminal equipment and protocols for telematic services. Recommendations T.0-T.63 (Study Group VIII).

FASCICLE VII.4 – Conformance testing procedures for the Teletex Recommendations. Recommendation T.64 (Study Group VIII).

FASCICLE VII.5 – Terminal equipment and protocols for telematic services. Recommendations T.65-T.101, T.150-T.390 (Study Group VIII).

FASCICLE VII.6 – Terminal equipment and protocols for telematic services. Recommendations T.400-T.418 (Study Group VIII).

FASCICLE VII.7 – Terminal equipment and protocols for telematic services. Recommendations T.431-T.564 (Study Group VIII).

Volume VIII

FASCICLE VIII.1 – Data communication over the telephone network. Series V Recommendations (Study Group XVII).

FASCICLE VIII.2 – Data communication networks: services and facilities, interfaces. Recommendations X.1-X.32 (Study Group VII).

FASCICLE VIII.3 – Data communication networks: transmission, signalling and switching, network aspects, maintenance and administrative arrangements. Recommendations X.40-X.181 (Study Group VII).

FASCICLE VIII.4 – Data communication networks: Open Systems Interconnection (OSI) – Model and notation, service definition. Recommendations X.200-X.219 (Study Group VII).

FASCICLE VIII.5 – Data communication networks: Open Systems Interconnection (OSI) – Protocol specifications, conformance testing. Recommendations X.220-X.290 (Study Group VII).

FASCICLE VIII.6 – Data communication networks: interworking between networks, mobile data transmission systems, internetwork management. Recommendations X.300-X.370 (Study Group VII).

FASCICLE VIII.7 – Data communication networks: message handling systems. Recommendations X.400-X.420 (Study Group VII).

FASCICLE VIII.8 – Data communication networks: directory. Recommendations X.500-X.521 (Study Group VII).

Volume IX – Protection against interference. Series K Recommendations (Study Group V). Construction, installation and protection of cable and other elements of outside plant. Series L Recommendations (Study Group VI).

Volume X

CONTENTS OF VOLUME V OF THE BLUE BOOK

Part I – Series P Recommendations

Telephone transmission quality

Part II – Supplements to Series P Recommendations

PRELIMINARY NOTES

1 This volume fully supersedes Volume V of the CCITT *Red Book* (Geneva, 1985).

2 The units used in this Volume are in conformity with CCITT Recommendations B.3 and B.4 (Volume I of the *Blue Book*).

The following abbreviations are used, particularly in diagrams and tables, and always have the following clearly defined meanings:

dBm the absolute power level in decibels;

dBm0 the absolute power level in decibels referred to a point of zero relative level;

dBr the relative power level in decibels;

dBm0p the absolute psophometric power level in decibels referred to a point of zero relative level.

The units for air pressure are related as follows:

1 Pa (Pascal) = 1 N (Newton)/m^2 = 10 dyne/cm^2 = 10 barye = 10 μbar.

3 In this Volume, the expression "Administration" is used for shortness to indicate both a telecommunication Administration and a recognized private operating agency.

4 Warning concerning some Recommendations:

A. In CCITT documents the designation "loudness rating" "LR", always refers to a specific CCITT definition. However, in technical literature and in some national specifications, the term "loudness rating" has ben used sometimes as a general concept and sometimes as a direct designation in loudness rating systems different from those defined by CCITT. Therefore, when CCITT LR values are cited in other contexts than in CCITT documents, care should be taken to avoid confusion. For instance, an introductory statement could be made such as: "All loudness rating designations refer to the CCITT definitions." Alternatively, "(CCITT)" could be added after the LR value. For example:

OLR = 10 dB (CCITT).

B. New technical developments as well as experience from existing telecommunication systems are reflected in CCITT work. Thus, new Recommendations have to be established and old Recommendations updated, sometimes at a fairly rapid rate, as a guidance to Administrations and manufacturers. Some organizational bodies, responsible for national and/or regional telecommunication specifications, use CCITT guidance in the form of specific references to particular CCITT Recommendations in their Standard Documents.

To avoid confusion and non-technical complications, such references to CCITT Recommendations should always be given together with a reference to the version of the CCITT Book in which the Recommendations are published. For example: "CCITT Rec. P. 79 (Blue Book)".

C. It should be observed that Supplements are only published for information purposes. Supplements do *not* have the status of agreed Recommendations.

PART I

Series P Recommendations

TELEPHONE TRANSMISSION QUALITY

PAGE INTENTIONALLY LEFT BLANK

PAGE LAISSEE EN BLANC INTENTIONNELLEMENT

SECTION 1

VOCABULARY
AND
EFFECTS OF TRANSMISSION PARAMETERS ON CUSTOMER OPINION OF TRANSMISSION QUALITY AND THEIR ASSESSMENT

Recommendation P.10

VOCABULARY OF TERMS ON TELEPHONE TRANSMISSION QUALITY AND TELEPHONE SETS

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

1 Introduction

This Recommendation contains terms and definitions appropriate to the work of Study Group XII which were discussed within the Group of Experts N of the Joint Coordinating Group for the CCIs and the IEC.

Terms which appear in the International Electrotechnical Vocabulary (IEV) (Chapter 722) have their IEV number reproduced here for reference purposes. Terms of the CCITT have been classified in a manner similar to that used in the IEV.

2 Terms and definitions

02. *Telephone set components*

02.01 **Y-ratio**

F: rapport Y

S: relación Y

The ratio between the sending and receiving efficiencies of a passive *telephone set* circuit.

04. *Telephone set types*

04.01 **telephone set; telephone instrument**

F: poste téléphonique: appareil téléphonique; téléphone

S: aparato telefónico; teléfono

An assembly of apparatus for *telephony* including at least a *telephone transmitter*, a *telephone receiver* and the wiring and components immediately associated with these transducers.

Note – A telephone set usually includes other components such as a *switchhook*, a built-in *telephone bell*, and a *dial*.

722.04.01

04.02 **telephone station**

F: poste téléphonique (installé)

S: estación telefónica

A *telephone set* with associated wiring and auxiliary equipment connected to a *telephone network* for the purpose of *telephony.*

Note – The auxiliary equipment may include, for example, an external *call indicating device,* a protector, a *local battery.*

722.04.02

04.03 **loudspeaking (telephone) set**

F: poste (téléphonique) à écoute (ou réception) amplifiée sur haut-parleur

S: aparato telefónico con altavoz; teléfono de altavoz

A handset *telephone* using a *loudspeaker* associated with an amplifier as a *telephone receiver.*

722.04.10

04.04 **hands free (telephone) set**

F: poste (téléphonique) mains-libres

S: aparato telefónico manos libres; teléfono manos libres

A *telephone set* using a loudspeaker associated with an amplifier as a telephone receiver and which may be used without a handset.

722.04.11

04.05 **group-audio terminals**

F: terminal audio de communication de groupe

S: terminal audio de groupo

A hands free set primarily designed for use by several users.

05. *Telephone set accessories*

05.01 **acoustic shock suppressor (in telephony)**

F: anti-choc (en téléphonie)

S: supresor de choques acústicos; antichoque (en telefonía)

A device associated with a *telephone station* and intended to prevent *acoustic shocks,* by setting an upper limit to the absolute values of the instantaneous electrical voltage that can be applied to the *telephone earphone.*

722.05.07

13. *Private telephone systems*

13.01 **private (telephone) installation**

F: installation (téléphonique) intérieure

S: instalación telefónica privada

A *telephone network* installed on the premises of a single individual or organization.

Note – By convention, private telephone installations include sets of *telephone stations* which are connected to one *subscriber's line*

722.13.01

21. *Telephone calls description*

21.01 **call attempt (by a user)**

F: (tentative d')appel (par un usager)

S: (tentativa de) llamada (por un usuario)

A sequence of operations made by a user of a telecommunication network trying to obtain the desired user or service.

Associated term: to *call*

722.21.01; identical to 701.03.04

21.02 **connection**

F: chaîne de connexion

S: cadena de conexión; conexión

A temporary association of transmission channels or telecommunication circuits, switching and other functional units set up to provide the means of a transfer of information between two or more points in a telecommunication network.

722.21.02; identical to 701.03.01

21.03 **(complete) connection**

F: chaîne de connexion complète; (chemin de) communication

S: cadena de conexión completa; conexión completa

A *connection* between users' terminals.

722.21.03; identical to 701.03.02

21.04 **call**

F: communication

S: comunicación

The establishment and use of a *complete connection* following a *call attempt*

722.21.04; identical to 701.03.05

31. *Local line networks*

31.01 **local line network**

F: réseau local de lignes (téléphoniques)

S: red local de líneas (telefónicas)

All the *subscribers' telephone lines* and ancillary equipment provided to connect *subscribers* to their *local switching entity.*

722.31.01

31.02 **subscriber's (telephone) line; subscriber loop (in telephony)**

F: ligne (téléphonique) d'abonné; ligne (de) réseau

S: línea (telefónica) de abonado; bucle de abonado (en telefonía)

A link between a public *switching entity* and a *telephone station* or a *private telephone installation* or another terminal using signals compatible with the *telephone network.*

Note – In French, the term "ligne de réseau" is used only when the private telephone installation is a *private branch exchange* or an *internal telephone system.*

722.31.02

31.03 **local (telephone) system; local (telephone) circuit**

F: système (téléphonique) local; circuit (téléphonique) local

S: sistema (telefónico) local

The combination of *subscriber's station, subscriber's line* and *feeding bridge* if present.

Note 1 – This term is used in the context of *transmission* planning and *performance.*

Note 2 – In CCITT English texts, the term "local (telephone) system" is preferred.

722.42.16

31.04 **subscriber system (in transmission planning)**

F: système d'abonné

S: sistema de abonado

A *subscriber's line* associated with that part of the *private telephone installation* connected to this line during a telephone *call.*

Note – This term is used in the context of *transmission* planning and *performance.*

722.42.17

32. *Telephone station usage*

32.01 **acoustic hood**

F: abri téléphonique; abriphone

S: cabina acústica; burbuja acústica

A hood lined with sound-absorbing material to facilitate the use of a *telephone station* by reducing the *ambient noise* level.

722.32.03

32.02 **telephone booth**

F: cabine téléphonique

S: cabina telefónica cerrada

A small cabin containing a *telephone station* and providing a certain measure of acoustic insulation and privacy for the user.

722.32.04

32.03 **telephone stall**

F: cabine téléphonique ouverte

S: cabina telefónica abierta

A *telephone booth* without a door.

722.32.05

41. *Transmission performance*

41.01 **acoustic shock (in telephony)**

F: choc acoustique (en téléphonie)

S: choque acústico (en telefonía)

Any temporary or permanent disturbance of the functioning of the ear, or of the nervous system, which may be caused to the user of a *telephone earphone* by a sudden sharp rise in the acoustic pressure produced by it.

Note – An acoustic shock usually results from the occurrence, in abnormal circumstances, of short-lived high voltages at the terminals of a *telephone set.*

722.41.20

41.02 **opinion score (in telephony)**

F: note d'opinion (en téléphonie)

S: nota de opinión (en telefonía)

The value on a predefined scale that a subject assigns to his opinion of the performance of the telephone transmission system used either for conversation or only for listening to spoken material.

Note – According to the IEV, the scale generally consists of five values, for example: excellent, good, fair, bad, unfair. This example does not correspond to CCITT practice (see Notes 2 and 3 of Recommendation P.82).

722.41.24

42. *Measuring apparatus*

42.01 **acoustic coupler (in telephonometry)**

F: coupleur acoustique (en téléphonométrie)

S: acoplador acústico (en telefonometría)

A cavity of defined shape and volume used for the testing of *telephone earphones* or *telephone transmitters* in conjunction with a calibrated microphone adapted to measure the pressure developed within the cavity.

722.42.12

42.02 **artificial ear**

F: oreille artificielle

S: oído artificial

A device for the calibration of earphones incorporating an *acoustic coupler* and a calibrated microphone for the measurement of sound pressure and having an overall acoustic impedance similar to that of the average human ear over a given frequency band.

722.42.13

42.03 **artificial mouth**

F: bouche artificielle

S: boca artificial

A device consisting of a *loudspeaker* mounted in an enclosure and having a directivity and radiation pattern similar to those of the average human mouth.

722.42.14

42.04 **artificial voice**

F: voix artificielle

S: voz artificial

A mathematically defined signal which reproduces human speech characteristics, relevant to the characterisation of linear and nonlinear telecommunication systems. It is intended to give a satisfactory correlation between objective measurements and tests with real speech.

42.05 **artificial voice**

F: voix artificielle

S: voz artificial

A complex sound, usually emitted by an artificial mouth and having a power sound spectrum corresponding to that of the average human voice.

722.42.15

42.06 **electrical artificial voice**

F: voix artificielle électrique

S: voz artificial eléctrica

The artificial voice produced as an electric signal, for testing transmission channels or other electric devices.

42.07 **acoustic artificial voice**

F: voix artificielle acoustique

S: voz artificial acústica

Acoustic signal at the MRP (Mouth Reference Point) of the artificial mouth. It complies with the same time and spectral specifications as the electrical artificial voice.

42.08 **artificial mouth excitation signal**

F: signal d'excitation de la bouche artificielle

S: señal de excitación de la boca artificial

A signal applied to the artificial mouth in order to produce the acoustic artificial voice. It is obtained by equalizing the electrical artificial voice for compensating the sensitivity/frequency characteristic of the mouth.

42.09 **head and torso simulator (HATS)**

F: simulateur de tête et de torse (STET)

S: simulador de cabera y tronco (SCT)

Manikin extending downward from the top of the head to the waist, designed to simulate the acoustic diffraction produced by a median adult and to reproduce the acoustic field generated by the human mouth.

43. *Telephonometry*

43.01 **reference equivalent**

F: équivalent de référence

S: equivalente de referencia

The loss, expressed in decibels, constant at all frequencies transmitted, which has to be introduced into the new *fundamental system for the determination of reference equivalents* or NOSFER in order to obtain in a given direction the same *loudness* as the *complete telephone connection* being considered, the *acoustical speech power* emitted by the talker being the same in both cases.

Note 1 – The reference equivalent is positive or negative according to whether it has been necessary for a loss to be added or removed from the NOSFER.

Note 2 – The reference equivalent is strictly defined by the measuring method described in Recommendation P.72 *(Red Book).*

722.43.14

43.02 **corrected reference equivalents**

F: équivalents de référence corrigés (ERC)

S: equivalentes de referencia corregidos (ERC)

Values of sending or receiving *reference equivalent* converted by a defined, nonlinear, transformation into corresponding values that obey the laws of algebraic addition.

Note – The conversion is performed to avoid some of the difficulties experienced in applying *reference equivalents.* It is defined in Annex C to Recommendation G.111.

722.43.17

43.03 **loudness rating**

F: équivalent pour la sonie

S: índice de sonoridad

A measure, expressed in decibels, for characterizing the *loudness* performance of *complete telephone connections* or of parts thereof such as *sending system, line, receiving system.*

Note – (added by the CCITT) – This definition is very general and corresponds to what is described as *loudness loss* in CCITT texts; in those texts, the term "loudness rating" should be confined to measurements in conformity with Recommendation P.76, and may be abbreviated as LR.

722.43.25

43.04 **R25 equivalent**

F: équivalent R25

S: equivalente R25

Loudness loss determined as a *reference equivalent* in accordance with Recommendation P.72 *(Red Book),* except that the listening level is constant, corresponding to 25 dB in NOSFER.

43.05 **planning equivalent**

F: équivalent de planification

S: equivalente de planificación

Result of a measurement with an objective meter which may be considered equal to an *R25 equivalent* or to a *corrected reference equivalent* with an accuracy which is sufficient for planning purposes.

43.06 **band sensation level**

F: niveau de sensation dans la bande

S: nivel de sensación en la banda

Difference, expressed in decibels, between the sound integrated over a frequency band and the sound pressure level in that band at the threshold of audibility, there being no other disturbing sound.

43.07 **earcap reference plane**

F: plan de référence écouteur

S: plano de referencia auricular

That plane formed by the contacting points of a flat surface against a telephone earcap.

43.08 **earcap reference point (ECRP)**

F: point de référence écouteur (PRE)

S: punto de referencia auricular (PRA)

Point in the *earcap reference plane,* used as a reference parameter.

43.09 **ear reference point (ERP)**

F: point de référence oreille (PRO)

S: punto de referencia oído (PRO)

A point located at the entrance to the ear canal of the listener's ear. (See figure A-1/P.64).

43.10 **earphone coupling loss (L_E)**

F: affaiblissement de couplage de l'écouteur (L_E)

S: pérdida de acoplamiento del auricular (L_E)

That quantity defined as the receiving sensitivity of a handset (usually as a function of frequency) when applied to an artificial ear minus the receiving sensitivity of the same handset on a human ear.

43.11 Δ_{SM} **(DELSM)**

F: Δ_{SM} (DELSM)

S: Δ_{SM} (DELSM)

Delta $_{SM}$ is defined as the difference between the sending sensitivity of a telephone set using a real mouth and voice, S_{MJ}, and that using a diffuse room noise source $S_{MJ/RN}$, such that:

$$\Delta_{SM} = S_{MJ/RN} - S_{MJ} \text{ dB.}$$

(See also Recommendations P.11, P.64, P.76, P.79, Supplement No. 11 and the Handbook on Telephonometry.)

Note – For most practical purposes Δ_{SM} will be closely approximated by the quantity Δ_{Sm} which is easier to determine.

43.12 Δ_{Sm} **(DELSm)**

F: Δ_{Sm} (DELSm)

S: Δ_{Sm} (DELSm)

Delta $_{Sm}$ is defined as the difference between the sending sensitivity of a telephone set using an artifical mouth S_{mJ}, and that using a diffuse room noise source $S_{mJ/RN}$, such that:

$$\Delta_{SM} = S_{MJ/RN} - S_{mJ} \text{ dB.}$$

(See also Recommendations P.11, P.64, P.76, P.79, Supplement No. 11 and the Handbook on Telephonometry.)

43.13 **lip plane**

F: position équivalente des lèvres

S: posición equivalente de los labios

Outer plane of the lip ring.

43.14 **lip ring**

F: anneau de garde (pour les lèvres)

S: anillo de labios

Circular ring of thin rigid rod, used for localizing the equivalent lip position of artificial mouths.

43.15 **guard-ring**

F: anneau de garde

S: anillo de guarda

Annular ring fitted, during tests, onto the transmitter housing of a telephone handset, to localize the sound source in a prescribed position relative to the microphone.

43.16 **metre air path**

F: trajet d'un mètre à l'air libre

S: trayecto de un metro en el aire

Measured reference of sound pressure loss over a 1 metre air path. In an anechoic environment, the sound pressure attenuation of such a path is approximately 30 dB measured from the MRP.

43.17 **modal distance**

F: distance modale

S: distancia modal

Distance between the centre of the microphone protective grid or front sound opening on a handset, and the centre of the guard-ring.

43.18 **modal gauge**

F: jauge modale

S: calibre modal

Template used to check a guard-ring position on a handset relative to the receiver *earcap reference plane.*

43.19 **modal position**

F: position modale

S: posición modal

Prescribed position and inclination of a handset relative to a fixed sound source.

43.20 **mouth reference point (MRP)**

F: point de référence bouche (PRB)

S: punto de referencia boca (PRB)

Point 25 mm in front of and on the axis of the lip position of a typical human mouth (or artificial mouth) (see Figure A-1/P.64).

43.21 **zero sidetone line impedance (Z_{S0})**

F: impédance de ligne à effet local nul

S: impedancia de línea de efecto local nulo (Z_{S0})

That circuit impedance which, when connected across the terminals of a telephone set, causes the sidetone to be reduced to zero.

43.22 **occlusion effect**

F: effet d'occlusion

S: efecto de oclusión

The change in human sidetone that occurs when the ear canal is occluded, e.g. by a telephone receiver.

43.23 **obstacle effect (obstruction effect)**

F: effet d'obstacle; effet d'obstruction

S: efecto de obstáculo; efecto de obstrucción

The change in the acoustic field close to a human or artificial mouth as obstacles (e.g. telephone transmitter) are brought into close proximity.

43.24 **sidetone path**

F: trajet d'effet local

S: trayecto de efecto local

Any path, acoustic, mechanical or electrical by which a telephone user's speech and/or room noise is heard in his own ear(s) (at ERP).

43.25 **sidetone path loss**

F: affaiblissement du trajet d'effet local

S: atenuación del trayecto de efecto local

The loss of the sidetone path expressed as a loss compared with the speech at the MRP. Symbols in common use are:

L_{MEHS} for sidetone paths within a human head,

L_{MEST} for electro-acoustic sidetone paths within the telephone set,

L_{MEMS} for mechanical sidetone paths within a telephone handset.

L_{RNST} for elctro-acoustic sidetone path from a diffuse room noise source to the earphone.

Each of these paths may be measured as sensitivities, in which case they become S_{MEHS}, S_{MEST}, S_{MEMS} and S_{RNST}, and experience a change of sign. Thus, for example, $S_{MEST} = -L_{MEST}$.

43.26 **listener sidetone rating (LSTR)**

F: affaiblissement d'effet local pour la personne qui écoute (AELE)

S: índice de efecto local para el oyente (IELO)

The loudness of a diffuse room noise source as heard at the subscriber's (earphone) ear via the electric sidetone path in the telephone instrument, compared with the loudness of the intermediate reference system (IRS) overall, in which the comparison is made incorporating a speech signal heard via the human sidetone path (L_{MEHS}) as a masking threshold.

43.27 **sidetone balance network**

F: réseau d'équilibrage d'effet local

S: red equilibradora del efecto local

An electrical network as part of a 2- to 4-wire balance point within a telephone set circuit for the purpose of controlling the telephone sidetone path loss.

43.28 **sidetone masking rating (STMR)**

F: affaiblissement d'effet local par la méthode de masquage (AELM)

S: índice de enmascaramiento para el efecto local (IEEL)

The loudness of a telephone sidetone path compared with the loudness of the intermediate reference system (IRS) overall in which the comparison is made incorporating the speech signal heard via the human sidetone path L_{MEHS} as a masking threshold.

43.29 **speech volume penalty**

F: pénalisation en volume sonore

S: penalización en volumen sonoro

The reduction in a subscriber's talking level (usually expressed as a function of a speech sidetone rating, e.g. STMR) due to the presence of sidetone.

43.30 **talking resistance**

F: résistance de conversation

S: resistencia de conversación

Fixed resistance used for test purposes, which has a resistance equal to that of a carbon microphone at a particular current.

43.31 **virtual source position**

F: position de la source virtuelle

S: posición de la fuente virtual

That position within a human or artificial mouth at which emitted sounds appear to have their source.

43.32 **virtual source function**

F: fonction de source virtuelle

S: función de la fuente virtual

The change in virtual source position as a function of some other parameter, e.g. frequency, proximity of obstacles.

43.33 **orthotelephonic reference condition**

F: condition de référence orthotéléphonique

S: condición de referencia ortotelefónica

Acoustic path between a talker and a listener, facing each other at a distance of 1 meter in the free field.

43.34 **orthotelephonic acoustic reference gain**

F: gain de référence acoustique orthotéléphonique

S: ganancia de referencia acústica ortotelefónica

Ratio of the pressure at the ear reference point of the listener to the pressure at the mouth reference point of the talker under othotelephonic reference conditions.

43.35 **total electroacoustic gain**

F: gain électroacoustique total

S: ganancia electroacústica total

Ratio of the pressure at the ear reference point of a listener to the pressure at the mouth reference point of a talker connected by a telephone channel.

43.36 **insertion gain (orthotelephonically referred gain)**

F: gain d'insertion (gain de référence orthotéléphonique

S: ganancia de inserción (ganancia referida ortotelefónicamente)

Ratio of the total electroacoustic gain to the orthotelephonic acoustic reference gain.

44. *Speech level measurements*

44.01 **active time**

F: durée d'activité

S: tiempo activo

Aggregate of all intervals of time when speech is deemed to be present according to the criterion adopted by CCITT (Recommendation P.56) for the purpose of measuring.

44.02 **active speech level**

F: niveau de conversation active

S: nivel vocal activo

A quantity, expressed in decibels relative to a stated reference, e.g. volts or pascals formed by averaging the speech-signal's power over the active time.

44.03 **activity factor**

F: coefficient d'activité

S: factor de actividad

Ratio of the active time to total timed elapsed during a measurement, usually expressed as a percentage.

44.04 **volume or speech volume**

F: volume ou volume de la parole

S: volumen ó volumen vocal

A quantity which is related to speech power and is measured at a stated point in a telephone circuit by means of a specified instrument, suitable for rapid real-time control or adjustment of level by a human observer (e.g. vu meter, ARAEN volume meter, peak programme meter).

44.05 **speech level**

F: niveau vocal

S: nivel vocal

A general term embracing speech volume, active speech level and any other similar quantity expressed in decibels relative to a stated reference.

Recommendation P.11

EFFECT OF TRANSMISSION IMPAIRMENTS

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

1 Purpose

An essential purpose of the present transmission plan for international connections is to provide guidance on the control of transmission performance. Such guidance is contained in Recommendations related to complete connections and to the constituent parts of a connection. These Recommendations contain performance objectives, design objectives and maintenance objectives, as defined in Recommendation G.102 for various transmission impairments which affect the transmission quality and customer opinion of transmission quality[1] Typical transmission impairments include transmission loss, circuit noise, talker echo, sidetone loss, attenuation distortion, group-delay distortion and quantizing distortion. Although not under the control of the transmission planner, room noise is another important factor which should be considered.

This Recommendation is concerned with the effect of transmission parameters, such as those listed above, on customer opinion of transmission quality. It is based on information contributed in response to specific questions which have been studied by the CCITT. Much of this information is based on the results of subjective tests in which participants have talked, listened or conversed over telephone connections with controlled or known levels of the impairments and rated the tansmission quality on an appropriate scale. General guidance for the conduct of such tests is provided in Recommendation P.80. In addition, Recommendation P.82 provides guidance on the use of telephone user surveys to assess speech quality on international calls.

Specific purposes of this Recommendation are:

1) to provide a general, but concise, summary of the major transmission impairments and their effect on transmission quality which would serve as a central reference for transmission planners;

2) to provide for retention of basic information on transmission quality in support of relevant Series P and Series G Recommendations with appropriate reference to these Recommendations and other sources of information such as Supplements and Questions under study;

3) to provide for the interim retention of basic information on transmission quality which is expected to be relevant in the formulation of future Recommendations.

---

[1] In this Recommendation, the term "impairment" is used in a general sense to refer to any characteristic or degradation in the transmission path which may reduce the performance or quality. It is not used to denote "equivalent loss" as was the case in some earlier CCITT texts.

§ 2 of this Recommendation provides a brief description of individual impairments which can occur in telephone connections, typical methods of characterization and general guidance on the acceptable levels of these impairments. More specific information is provided in Annexes to this Recommendation, in other Recommendations and in Supplements.

§ 3 of this Recommendation is concerned with the effect of combined impairments on transmission quality and the use of opinion models which permit estimates to be made of customer opinion as a function of combinations of transmission impairments in a telephone connection. Thus, they can be used to evaluate the transmission quality provided by the present transmission plan, the impact of possible changes in the transmission plan or the consequences of departures from the transmission plan. Such evaluations require certain assumptions concerning the constituent parts of a connection, and guidance is provided by the hypothetical reference connections which are the subject of Recommendations G.103 and G.104.

2 Effect of individual impairments

2.1 *General*

§ 2 describes invididually a number of the transmission impairments which can affect the quality of speech transmission in telephone connections. Information is provided on the general nature of each impairment, on methods which have been recommended to measure the impairment and on the acceptable ranges for the impairment. References are provided to Recommendations where more detailed information on measurement methods and recommended values can be found.

2.2 *Loudness loss*

An essential purpose of a telephone connection is to provide a transmission path for speech between a talker's mouth and the ear of a listener. The loudness of the received speech signal depends on acoustic pressure provided by the talker and the loudness loss of the acoustic-to-acoustic path from the input to a telephone microphone at one end of the connection to the output of a telephone receiver at the other end of the connection. The effectiveness of speech communication over telephone connections and customer satisfaction depend, to a large extent, on the loudness loss which is provided. As the loudness loss is increased from a preferred range, the listening effort is increased and customer satisfaction decreases. At still higher value of loudness loss, the intelligibility decreases and it takes longer to convey a given quantity of information. On the other hand, if too little loudness loss is provided, customer satisfaction is decreased because the received speech is too loud.

Over the years, various methods have been used by transmission engineers to measure and express the loudness loss of telephone connections. The reference equivalent method is a subjective method which has been widely used in CCITT and is defined in Recommendations P.42 and P.72 *(Red Book).*

Because difficulties were encountered in the use of reference equivalents, the planning value of the overall reference equivalent was replaced by the corrected reference equivalent (CRE) as defined in Recommendation G.111 (CCITT *Red Book*). This change required some adjustment in the recommended values of loudness loss for complete and partial connections.

Recommendations P.76, P.78 and P.79 provide information on subjective and objective methods for the determination of loudness ratings (LRs) which are now recommended. These methods are expected to eliminate the need for the subjective determinations of loudness loss in terms of the corrected reference equivalent. The currently recommended values of loudness loss in terms of loudness ratings are given in Recommendations G.111 and G.121.

2.2.1 *Customer opinion*

Customer opinion, as a function of loudness loss, can vary with the test group and the particular test design. The opinion results presented in Table 1/P.11 are representative of laboratory conversation test results for telephone connections in which other characteristics such as circuit noise are contributing little impairment. These results indicate the importance of loudness loss control.

TABLE 1/P.11

Overall loudness rating (dB)	Representative opinion results [a]	
	Percent "good plus excellent"	Percent "poor plus bad"
5 to 15	> 90	< 1
20	80	4
25	65	10
30	45	20

[a] Based on opinion relationship derived from the transmission quality index (see Annex A).

2.2.2 *Recommended values of loudness rating*

Table 2/P.11 provides further information on selected values of loudness rating which have been recommended or are under study by the CCITT.

Note – Recommended values of loudness ratings are under study in Question 19/XII.

2.3 *Circuit noise*

The circuit noise in a telephone connection has a major effect on customer satisfaction and the effectiveness of speech communication. This noise may include white circuit noise and intermodulation noise from transmission systems as well as hum and other types of interference such as impulse noise and single frequency tones. Customer satisfaction depends on the power, the frequency distribution and the amplitude distribution of the noise. For a given type of noise, the satisfaction generally decreases monotonically with increasing noise power.

Circuit noise is generally expressed in terms of the indications given by a psophometer standardized by the CCITT in Recommendation O.41. With this apparatus, frequency-weighted measurements of noise power in dBmp can be made a various points in telephone connections.

Note – Although the psophometer is normally used to measure wideband circuit noise, some subjective tests indicate that it satisfactorily characterizes the subject interfering effect of induced power hum on message circuits.

2.3.1 *Opinion results*

Many tests have been conducted which demonstrate the effect of circuit noise on customer opinion. These tests have shown that opinion judgements of circuit noise are also highly dependent on the loudness loss of the connection and can be influenced by many other factors, particularly the room noise and sidetone loss.

The subjective effect of circuit noise measured at a particular point in a telephone connection depends on the electrical-to-acoustical loss or gain from the point of measurement to the output of the telephone receiver. As a convenience in assessing the contributions from different sources, circuit noise is frequently referred to the input of a receiving system with a specified receiving CRE or loudness rating. A common reference point is the input of a receiving system having a Receiving CRE of 0 dB. When circuit noise is referred to this point, circuit noise values less than −65 dBmp have little effect on transmission quality in typical room noise environments. Transmission quality decreases with higher values of circuit noise.

The opinion results presented in Table 3/P.11 are representative of laboratory conversation tests and illustrate the effect of circuit noise when other connection characteristics such as loudness are introducing little additional impairment. When the loudness loss is greater than the preferred range, the effect of a given level of circuit noise becomes more severe.

Note – See Annex A of this Recommendation for further information on the effects of circuit noise.

TABLE 2a/P.11

Values (dB) of reference equivalent RE (q), and corrected reference equivalent CRE (y) for various connections cited in Red Book Recommendations G.111 and G.121 (send and receive interfaces are at the virtual analogue switching point, VASP)

		Previously recommended RE (q)	CRE (y)
Optimum range for a connection (Rec. G.111, § 3.2)	min optimum max	6 9 18	5 [a)] 7 [a)] to 11 16
Traffic weighted mean values			
Long term objectives			
– connection (Rec. G.111, § 3.2)	min max	13 18	13 16
– national system send (Rec. G.121, § 1)	min max	10 13	11.5 13
– national system receive (Rec. G.121, § 1)	min max	2.5 4.5	2.5 4
Short term objectives			
– connection (Rec. G.111, § 3.2)	max	23	25.5
– national system send (Rec. G.121, § 1)	max	16	19
– national system receive (Rec. G.121, § 1)	max	6.5	7.5
Maximum values for national system (Rec. G.121, § 2.1) of an average-sized country	send receive	21 12	25 14
Minimum for the national sending system (Rec. G.121, § 3)		6	7

a) These values apply for conditions free from echo; customers may prefer slightly larger values if some echo is present.

TABLE 2b/P.11

LR values as cited in Recommendations G.111 and G.121

	SLR [a]	CLR [a]	RLR [a]	OLR [a]
Traffic-weighted mean values:				
long term	7-9 [b]	0-0.5 [e]	1-3 [b) f]	8-12 [e) f) g]
short term	7-15 [b]	0-0.5 [e]	1-6 [b) f]	8-21 [e) f) g]
Maximum values for an average-sized country:	16.5 [c]		13 [c]	
Minimum value:	– 1.5 [d]			

a) As in Figure 1/P.11.

b) Rec. G.121, § 1.

c) Rec. G.121, § 2.1.

d) Rec. G.121, § 3.

e) When the international chain is digital, CLR = 0. If the international chain consists of one analogue circuit, CLR = 0.5, and then OLR is increased by 0.5 dB. (If the attenuation distortion with frequency of this circuit is pronounced, the CLR may increase by another 0.2 dB. See Annex A, § A.4.2 to Recommendation G.111.)

f) See also the remarks made in Rec. G.111, § 3.2.

g) Rec. G.111, § 3.2.

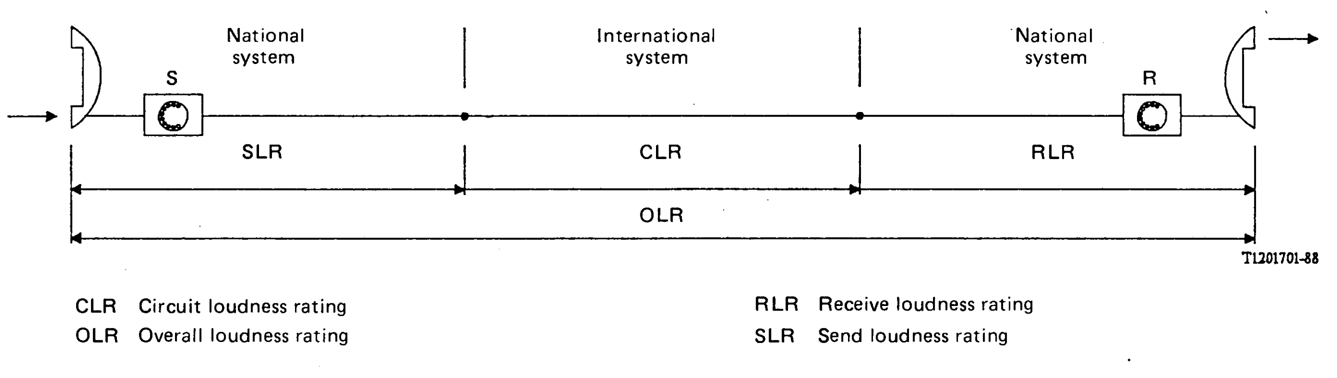


FIGURE 1/P.11

Designation of LRs in an international connection

TABLE 3/P.11

Circuit noise at point 0 dB RLR (dBmp)	Representative opinion results [a]	
	Percent "good plus excellent"	Percent "poor plus bad"
−65	> 90	< 1
−60	85	2
−55	75	5
−50	65	10
−45	45	20

[a] Based on opinion relationship derived from the transmission quality index (see Annex A).

2.3.2 *Recommended values of circuit noise*

Contributions to circuit noise from the various parts of a connection should be kept as low as practical. The major source of circuit noise on medium or long connections is likely to occur in analogue transmission facilities where the noise power is typically proportional to the circuit length. In Recommendation G.222, a noise objective of 10 000 pW0p or −50 dBm0p is recommended for the design of carrier transmission systems of 2500 km. When referred to a point of 0 dB receiving loudness rating (assuming a loss of 6 to 12 dB), this corresponds to a noise level in the range from −62 to −56 dBmp, which is sufficiently high to affect the transmission quality.

The decrease in quality is larger on longer circuits or in connections with several such circuits in tandem. The CCITT states in Recommendation G.143 that it is desirable 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. This corresponds to approximately −46 dBm0p at the end of the chain or −58 to −52 dBmp at a point with a 0 dB receiving reference equivalent. Other sources of circuit noise in international connections should be controlled such that their contribution is small compared to that permitted on analogue transmission facilities. Specific guidance is provided in a number of Recommendations.

The limits for a single tone or narrow bands of noise should be more stringent than the limits for wideband noise in order to avoid customer annoyance. As a general rule to limit annoyance from single frequency tones, the power in any individual tone should be 10 dB less than the psophometric noise power in the circuit. To avoid audibility, an additional 5 dB of margin is recommended where practical.

Note – The effect of impulse noise depends on the rate of occurrence. For pulses which were damped 2 kHz oscillatory transients with durations of about one millisecond (a pulse shape commonly encountered on message facilities), limited test results have been reported in terms of the mean value of the peak power of the individual impulses measured on the line at the telephone set. The results indicate that the noise pulses occurring at an average rate of one per second or less are not annoying if their mean intensity is less that 65 dBrn (−25 dBm). At the rate of 45 per second, an acceptable level of 30 dBrn (−60 dBm) was indicated.

2.4 *Sidetone*

Sidetone of a telephone set is the transmission of sound from the telephone microphone to the telephone receiver in the same telephone set. Thus, the sidetone path of a telephone set is one of the paths through which the talker hears himself as he speaks. Other such paths are the head conduction path and the acoustic path from the mouth to the ear through earcap leakage. The presence of these other paths affects the customer's perception of sidetone and consequently his reaction to it.

Sidetone affects telephone transmission quality in several ways. Too little sidetone loss causes the returned speech levels to be too loud and this reduces customer satisfaction. Another aspect of insufficient sidetone loss is that talkers tend to reduce their speech levels and/or move the handset away from the mouth, thus reducing the received levels at the far end of the connection. Handset movement can also reduce the seal at the ear and thus make it easier for room noise to reach the ear through the resulting leakage path, while reducing as well the level of the received signal from the far end of the connection. In addition, the sidetone path provides another route by which room noise can reach the ear. Very low levels of sidetone loss can effect transmission quality adversely. As the sidetone loss is increased there is a general region of preferred loss values. Excessive sidetone loss can make a telephone set sound dead as one is talking and, for many connections, the absence of sidetone would not be a preferred condition.

Sidetone loss has, in the past, been rated as a loudness loss in much the same manner as connection loudness loss, for example, in terms of sidetone reference equivalent (STRE) (Recommendation P.73, *Red Book*). A better method, which yields ratings that correlate with the subjective effects of sidetone, for a subscriber when considered as a talker, is described in Recommendation P.76. This method, Sidetone Masking Rating (STMR), takes into account the head conduction and direct acoustic paths as a masking threshold.

Recent studies have shown that, due to the increasing use of linear microphones in telephone handsets, a rating method is also necessary to control the loudness of room noise heard via the telephone sidetone path by means of a Listener Sidetone Rating (LSTR). LSTR (Recommendations P.76 and P.79) uses the same concept and calculation algorithm as STMR, but the sidetone sensitivity is measured using a room noise source rather than an artificial mouth source.

The sidetone loss is influenced by the telephone set design and the impedance match between the telephone set and the subscriber line. Impedance variations at the far end of the subscriber line can also have significant mismatch effects on short subscriber lines with low loss. Impedance mismatches at other points in the connection will also affect the returned signal, but, as the delay in the return path becomes significant, the effect is generally considered as talker echo (see § 2.9).

2.4.1 *Recommended values of sidetone loss*

Recommendation G.121, § 5 provides guidance on preferred sidetone levels under various connection conditions for the subscriber both as a talker (STMR) and listener (LSTR).

Subjective test results of customer opinion as a function of sidetone loss in terms of STMR indicate a preferred range of 7 to 12 dB (see also Supplement No. 11). Lower values cause a substantial reduction in subscriber opinion and should only be used with caution. High values, up to 20 dB are acceptable, but higher values cause the impression of a "dead" connection.

To control the effects of high level room noise, the value of LSTR to strive for 13 dB. In general, this will not always be possible as, for most telephone sets having linear microphones and speech circuits, LSTR is closely linked to, and typically 1.5 to 4 dB greater than, STMR. [The relationship is determined by Δ_{sm} (DELSM), the difference between the microphone sensitivity when measured with a room noise source and when measured with a mouth. See Recommendations P.64, P.10, P.79, Supplement No. 11 and Annex A to Recommendation G.111, § A.4.3.3.]

Thus, connections having low values of STMR will generally also exhibit low values of LSTR.

2.5 *Room noise*

Room noise is the term used to describe the background noise in the environment of the telephone set. In a residential location it may consist of household appliances, radio or phonograph noise, conversations or street noise. In an office location, business equipment, air conditioning equipment and conversations are likely to predominate. In many situations, the effect of room noise may be inconsequential compared to the effects of circuit noise. In noisy locations such as call offices in public places, however, the effects of room noise may have a substantial effect on the ease of carrying on a conversation or even in being able to hear and understand properly.

Room noise can manifest itself in several ways. One is through leakage around the earcap of the receiver. Another is through the sidetone path of the telephone set if the sidetone loss is sufficiently low in comparison with leakage past the earcap (see § 2.4 above). A third way is through the other ear, although the effect of this on telephone reception is usually less than that of noise entering the "telephone ear", unless the sound in the room causes distraction (a baby crying, for example). A fourth way is through the transmitter over the connection to the receiving telephone set.

The previous discussion applies primarily to conventional telephone sets. Loudspeaking telephone sets are more susceptible to room noise.

Noise present in stationary or moving vehicles (not commonly referred to as a room noise) may also have a substantial effect on the ease of carrying on a conversation or in being able to hear and understand properly over telephone connections involving mobile station.

2.6 *Attenuation distortion*

Attenuation distortion is characterized by transmission loss (or gain) at other frequencies relative to the transmission loss at 800 or 1000 Hz. Thus, attenuation distortion includes the low-frequency and high-frequency rolloffs which determine the effective bandwidth of a telephone connection, as well as in-band variations in loss as a function of frequency. The loudness loss and articulation of a telephone connection are respectively a function of the attenuation distortion. Even when the loudness loss is maintained at a constant value, opinions of the transmission quality as determined by subjective tests usually get worse as the amount of attenuation distortion increases.

The effect of attenuation distortion on loudness is greater at the lower end of the frequency band than at the higher end. The effect of attenuation distortion on sound articulation is, on the contrary, more marked at the higher frequencies. For both loudness and articulation impairments due to bandpass characteristics, it can be assumed that the impairment values due to highpass and lowpass characteristics add directly if each attenuation distortion slope is greater than 15 dB/octave.

The effect of attenuation distortion on listening and conversation opinion scores decreases noticeably as the overall loudness loss of a connection increases, particularly when circuit noise also exists. The effect of attenuation distortion on opinion scores is typically less than that of loudness loss, particularly at high values of loudness loss, but may be comparable to that of noise when the values of loudness loss and noise are both low.

The current network performance objectives for attenuation distortion in the electrical transmission elements of a worldwide 4-wire chain of 12 circuits are given in Recommendation G.132 but, of course, the frequency characteristics of the telephone sets themselves have some influence.

Note – Further information on the effects of attenuation distortion on transmission quality are provided in Annex B.

2.7 *Group-delay distortion*

Group-delay distortion is characterized by the group delay at other frequencies relative to the group delay at the frequency where the group delay has its minimum value. Although the effect of group-delay distortion is usually a more significant impairment for data transmission than for speech transmission, large amounts of group-delay distortion can cause noticeable distortion for speech signals.

The effect of group-delay distortion at the upper and lower edges of the transmitted band can be described as "ringing" and "speech blurred", respectively. In the absence of noise or attenuation distortion, the effect is conspicuous throughout the entire range of typical loudness loss values. However, the effect in a typical 4-wire circuit chain is usually not serious since the group-delay distortion is normally accompanied by closely related attenuation distortion which tends to reduce the effect.

The current performance objectives for group-delay distortion for a worldwide chain of 12 circuits are given in Recommendation G.133.

Note – Further information on the effect of group-delay distortion is provided in Annex C.

2.8 *Absolute delay*

Values of absolute delay typical of those present in terrestrial transmission facilities have little effect on speech transmission quality if there is no talker or listener echo (4-wire connections, for example) or if the talker and listener echo are adequately controlled. Satellite facilities introduce larger amounts of delay (approximately 300 ms in each direction of transmission) and, again, the available opinion data indicates that there is little effect on the transmission quality of connections with a single satellite circuit, provided talker and listener echo are adequately controlled. Less data are available on the effects of one-way delays of approximately 600 ms (two satellite circuits in tandem) and the results are not entirely consistent. Therefore, caution is recommended with regard to the introduction of one-way absolute delay significantly greater than 300 ms.

Note – The effects of echo, echo control and propagation time are under study in Question 27/XII.

2.9 *Talker echo*

Talker echo occurs when some portion of the talker's speech signal is returned with enough delay (typically more than about 30 ms) to make the signal distinguishable from normal sidetone. Talker echo may be caused by reflections at impedance mismatches or by other processes such as go-to-return crosstalk. The effect of talker echo is a function of the loss in the acoustic-to-acoustic echo path and the delay in the echo path. In general, customer satisfaction is decreased as the loss of the echo path is decreased or the delay of the echo path is increased.

The overall loudness rating of the echo path is here defined as the sum of:

- the loudness rating in the two directions of transmission of the local telephone system of the talking subscriber (assumed to have minimum values of loudness rating);
- the loudness rating in the two directions of transmission of the chain of circuits between the 2-wire end of the local telephone system of the talking subscriber and the 2-wire terminals of the 4W/2W terminating set at the listener's end;
- the mean value of the echo balance return loss at the listener's end.

Echo tolerance curves are provided in Figure 2/G.131 which indicate the recommended LR of the echo path to control the probability of objectionable echo.

Note – The effect of echo and propagation time is under study in Question 27/XII.

2.10 *Listener echo*

Listener echo refers to a transmission condition in which the main speech signal arrives at the listener's end of the connection accompanied by one or more delayed versions (echoes) of the signal. Such a condition can occur as the result of multiple reflections in the transmission path. A simple, yet common, source of listener echo is a low loss 4-wire transmission path which interconnects two 2-wire subscriber lines. In such a connection, reflections can occur as the result of impedance mismatch at the hybrids at each end of the 4-wire section. A portion of the main speech signal can thus be reflected at the far end of the 4-wire path, return to the near end and be reflected again. The result is a listener echo, whose magnitude, relative to the main signal, depends on the two return losses and the two-way loss or gain of the 4-wire transmission path. The delay of the echo is determined primarily by the two-way delay of the 4-wire transmission path. For small delays, the listener echo results in a change in the spectral quality of the speech. For longer delays, the echo is more pronounced and is sometimes referred to as a "rain barrel" effect.

Listener echo may be characterized by the additional loss and additional delay in the listener echo path relative to that in the main signal path. The minimum value of the additional listener echo path loss over the frequency band of interest provides a margin against instability or oscillation. As a result, listener echo is frequently referred to as near-singing distortion. Recommendation G.122 provides guidance on the influence of national networks on stability in international connections.

2.11 *Nonlinear distortion*

Nonlinear distortion, in its most general sense, occurs in systems in which the output is not linearly related to the input. A simple example is a system in which the output signal can be represented, as a function of the input signal $e_i(t)$, by a power series of the form:

$$e_o(t) = a_1 e_i(t) + a_2 e_i^2(t) + a_3 e_i^3(t) + \dots,$$

which, in the case of a sinusoidal input, creates second, third and higher order harmonics in the output signal. For more complex signals, the nonlinear terms are frequently referred to as intermodulation distortion. Nonlinear distortion is normally a more significant impairment for data transmission than it is for speech transmission, but it can also be important for speech.

Up until now, one of the major sources of nonlinear distortion in telephone connections has been telephone sets using carbon microphones. Although carbon microphones are now being rapidly replaced by linear microphones, additional potential sources of nonlinear distortion are being introduced, e.g. by the use of digital encoding schemes, especially at low bit-rates. Theses schemes introduce quantizing distortion (see § 2.12) which is a particular form of nonlinear distortion. In addition, other devices such as syllabic compandors and overloaded amplifiers may be significant contributors.

Further information relevant to carbon and linear microphones is provided in Annex D, while Annex F contains information on the subjective effects of nonlinear distortion in general.

Note – Nonlinear distortion (and especially the definition of a suitable objective measuring method) is being studied under Question 13/XII.

2.12 *Quantizing distortion*

Quantizing distortion occurs in digital systems when an analogue signal is sampled and each sample is encoded into one of a finite set of values. The difference between the original analogue signal and that which is recovered after quantizing is called quantizing distortion or quantizing noise. For many digital encoding algorithms, such as A-law or μ-law PCM, which have a nearly-logarithmic companding law, the subjective effect of quantizing distortion can be approximated by adding signal-correlated noise (white noise which has been modulated by the speech signal). Such a signal can be generated in a modulated-noise reference unit which can be adjusted to provide a reference signal with a selected and nearly constant signal to signal-correlated-noise ratio. Recommendation P.70 describes the modulated-noise reference unit recommended by CCITT for use in evaluating digital codecs for telephone speech applications. The signal to signal-correlated-noise ratio, when expressed in decibels, is called Q. The effective Q of an unknown digital system can be determined by subjective comparison with the modulated-noise reference unit. (Supplement No. 14 provides guidelines on use of the modulated noise reference unit of Recommendation P.81.)

Subjective test results have been reported by some Administrations which have evaluated the effects of both circuit noise and Q on customer opinion. Results from tests of this type permit estimates to be made of the circuit noise level, which could provide approximately the same transmission quality ratings as a given level of quantizing distortion.

Note – Further information is provided in Annex E. The transmission performance of digital systems is under study in Question 18/XII.

2.13 *Phase jitter*

Phase jitter occurs when the desired signal, during transmission, is phase- or frequency-modulated at a low-frequency rate. If such distortion is present in sufficient quantity, the transmission quality is degraded. Table 4/P.11 summarizes the threshold data for single-frequency phase jitter which have been reported by one Administration. The results are in terms of the mean threshold expressed in terms of the signal-to-first order-sideband (C/SB) ratio in decibels. The average standard deviation across subjects was about 4 dB.

2.14 *Intelligible crosstalk*

Intelligible crosstalk occurs when the speech signal from one telephone connection is coupled to another telephone connection such that the coupled signal is audible and intelligible to one or both of the participants on the second telephone connection. Although the level of the intelligible crosstalk may be high enough to degrade the transmission quality, the major concern is the loss of privacy.

TABLE 4/P.11

Phase jitter modulation rate (Hz)	Mean threshold C/SB ratio (dB)	
	Male talkers	Female talkers
25	10.9	13.8
80	14.4	16.3
115	12.3	18.3
140	13.8	20.0
200	17.0	18.0

A number of factors influence the intelligibility of a signal which is coupled from one telephone connection to another. They include the characteristics of the telephone apparatus (including sidetone), circuit noise, room noise, the coupling loss, the interfering talker's speech level and the hearing acuity of the listener.

Information is provided in Recommendation P.16 on the intelligibility threshold for crosstalk and on methods for calculating the probability of intelligible crosstalk. Design objectives for the various apparatus in telephone connections should be selected such that the probability of intelligible crosstalk is sufficiently low. Typically, objectives are intended to keep the probability below one percent in connections where the interfering and interfered-with parties are unlikely to know each other and unlikely to suffer the same coupling again. A more stringent objective of 0.1 percent is typical for use in local equipment such as subscriber lines where the two parties may be neighbours.

3 Effect of multiple impairments and the use of opinion models

Transmission performance of a practical connection can be affected by several transmission impairments which are likely to coexist. Although results for customer opinion in the form described in § 2 are useful in many studies involving one or two types of transmission impairments, they become increasingly cumbersome as the number of impairments under study increases. This has led to the study of more extensive analytical models of customer opinion which can be based on the composite results of a number of individual tests and studies. The formulation and use of these more comprehensive models are aided by the availability of modern digital computers. Ideally, such models might eventually include the effects of all or most of the significant types of transmission impairment mentioned in § 2 above.

Note – Although some Administrations have reported on efforts directed toward this goal, the subject of models for predicting transmission quality from objective measurements is still under study in Question 7/XII [11]. Examples of opinion models used by Bellcore, British Telecom, NTT and CNET are given in Supplement No. 3 at the end of this Volume.

ANNEX A

(to Recommendation P.11)

Transmission quality index

A.1 *Introduction*

This annex which was prepared as part of the reply to Question 4/XII (1985-1988) describes a simple conversation opinion model for predicting the combined effects of overall loudness rating (OLR) in terms of Recommendation P.79 and psophometric noise in dBmp. It also includes the efforts of sidetone masking rating (STMR), room noise in dBA and attenuation distortion.

A.2 *Connection parameters used in the model*

The following list gives the connection parameters and their range of values.

	Connection parameters	*Range*
OLR	Overall loudness rating in dB	0 to 40
CN	Circuit noise at 0 dB, RLR in dBmp	−80 to −40
RN	Room noise in dBA	30 to 70
Q	Signal/quantizing distortion in dB	0 to 100
STMR(T)	Sidetone masking rating (talker end) in dB	0 to 20
STMR(L)	Sidetone masking rating (listener end) in dB	0 to 20
FL	Lower cutoff frequency (10 dB) in Hz	200 to 600
FU	Upper cutoff frequency (10 dB) in Hz	2500 to 3400

A.3 *Basic model for transmission quality index*

$I = I(S/N)I(BW)I(ST)$ (A-1)

$I(S/N)$ = Index for loudness loss and circuit noise

$= 1.026 - 0.013\sqrt{(OLRe - OLRp)^2 + 4} - 0.01(NT + 80)$ (A-2)

$OLRe$ = Effective OLR with effect of STMR(T) on speech level

$= OLR$ for $STMR(T) > 12$ dB

$= OLR + [12 - STMR(T)]/3$ for $STMR(T) < 12$ dB (A-3)

$OLRp$ = Optimum value of OLR as function of CN and RN

$= 10 - (NT + 80)/10$ (A-4)

NT = Circuit noise equivalent of all noise in dBmp

$= N1\ (+)\ NF\ (+)\ N(Q)$ (A-5)

$N1$ = Circuit noise equivalent of circuit noise and room noise in dBmp

$= CN\ (+)\ RNE(L)\ (+)\ RNE(S)$ (A-6)

$RNE(L)$ = Circuit noise equivalent due to room noise and earcap leak in dBmp

$= RN - 116$ (A-7)

$RNE(S)$ = Circuit noise equivalent due to room noise and sidetone path in dBmp

$= RN - 100 - STMR(L) - D$ (A-8)

D = Sidetone rating for room noise − $STMR(L)$

$= 15 - 0.006\,(RN - 30)^2$ (Carbon Transmitter) (A-9)

$= 3$ (Linear Transmitter)

NF = Apparent noise floor

$= -70$ dBmp (default value) (A-10)

NQ = Circuit noise equivalent of quantizing distortion in dBmp

$= -3 - OLR - 2.2Q$ (A-11)

$I(BW)$ = Index for bandwidth

$= [1 - 0.0008(FL - 200)]\,[1 - 0.00022(3400 - FU)]$ (A-12)

$I(ST)$ = Index for sidetone

$= 1 - 0.00003(OLRe)\,[STMR(L) - 15]^2$ (A-13)

$$FI = 7.2I - 2 \quad \text{(A-14)}$$

$$X = 0.96(FI - 2) + 0.041(FI - 2)^3 \quad \text{(A-15)}$$

$$MOS = 4 \exp(X)/[1 + \mathrm{EXP}(X)] \quad \text{(A-16)}$$

$$\%(G + E) = 100/[1 + \exp(-QA)] \quad \text{(A-17)}$$

$$QA = 1.59577\, A\, (1 + 0.04592\, A^2 - 0.000368\, A^4 + 0.000001\, A^6) \quad \text{(A-18)}$$

$$A = FI - 2.5 \quad \text{(A-19)}$$

$$\%(P + B) = 100 - 100/[1 + \exp(-QB)] \quad \text{(A-20)}$$

$$QB = 1.59577\, B\, (1 + 0.04592\, B^2 - 0.000368\, B^4 + 0.000001\, B^6) \quad \text{(A-21)}$$

$$B = FI - 1.5 \quad \text{(A-22)}$$

G = Good

P = Poor

E = Excellent

B = Bad

A.4 *Typical results*

Typical results from the model in terms of mean opinion score (MOS) are shown in Figures A-1/P.11 to A-7/P.11.

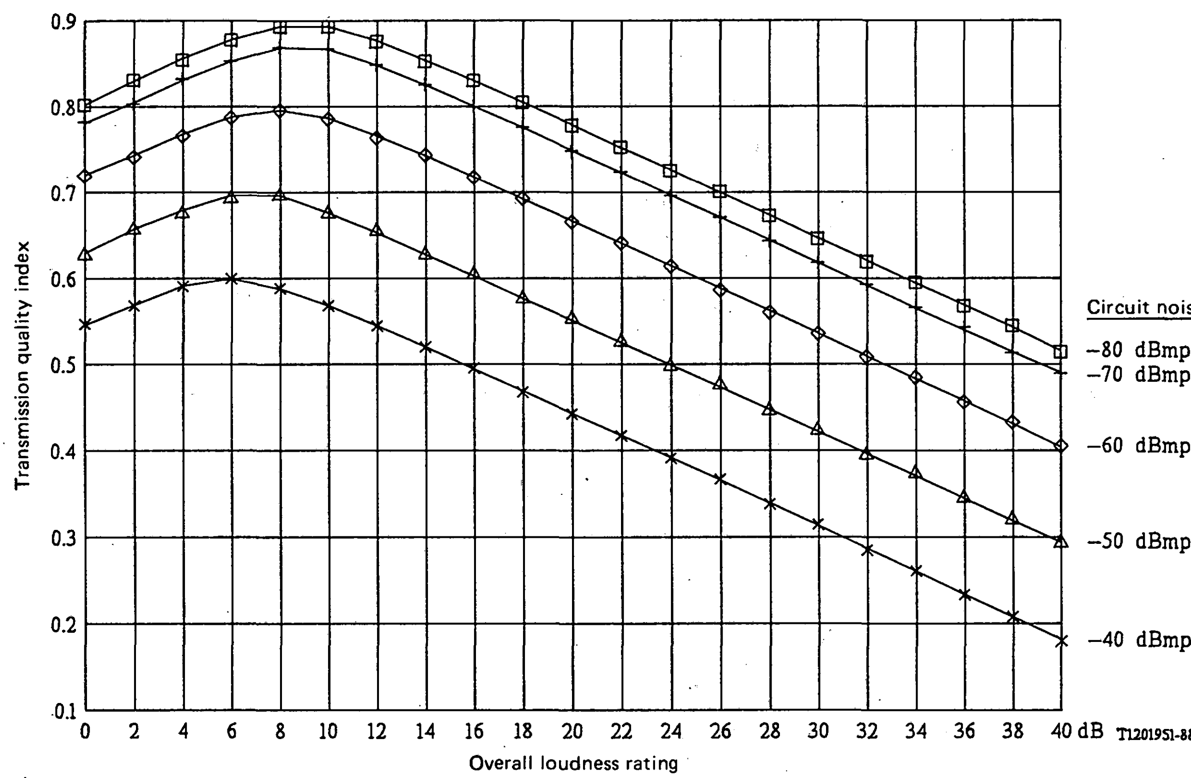


FIGURE A-1/P.11

Transmission quality index as a function of overall loudness rating and circuit noise

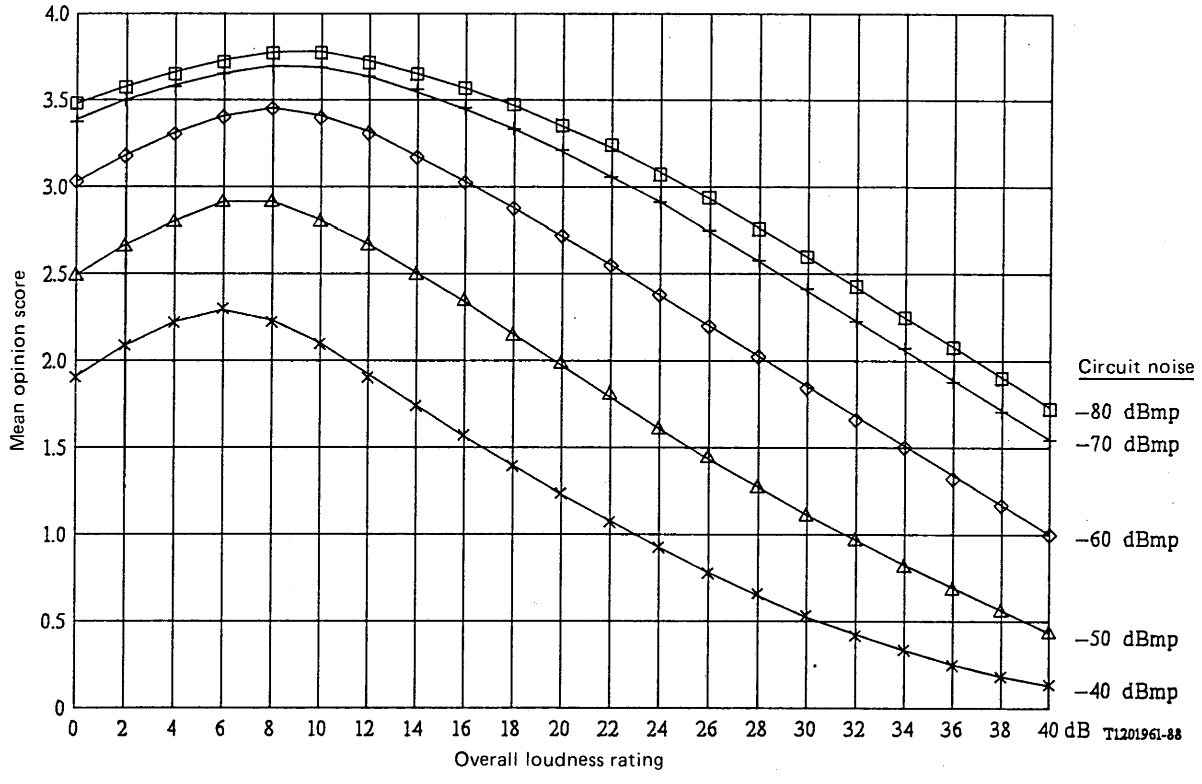


FIGURE A-2/P.11

Mean opinion score as a function of overall loudness rating and circuit noise

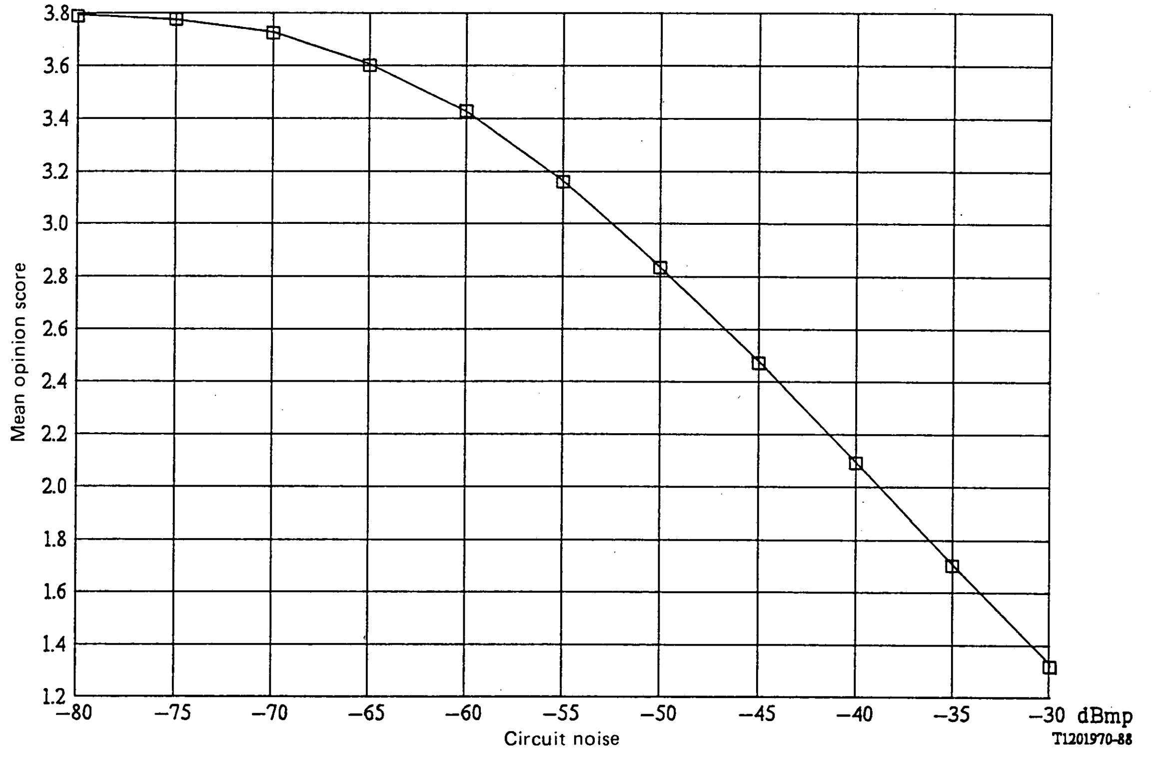


FIGURE A-3/P.11

Mean opinion score as a function of circuit noise with OLR = 10 dB

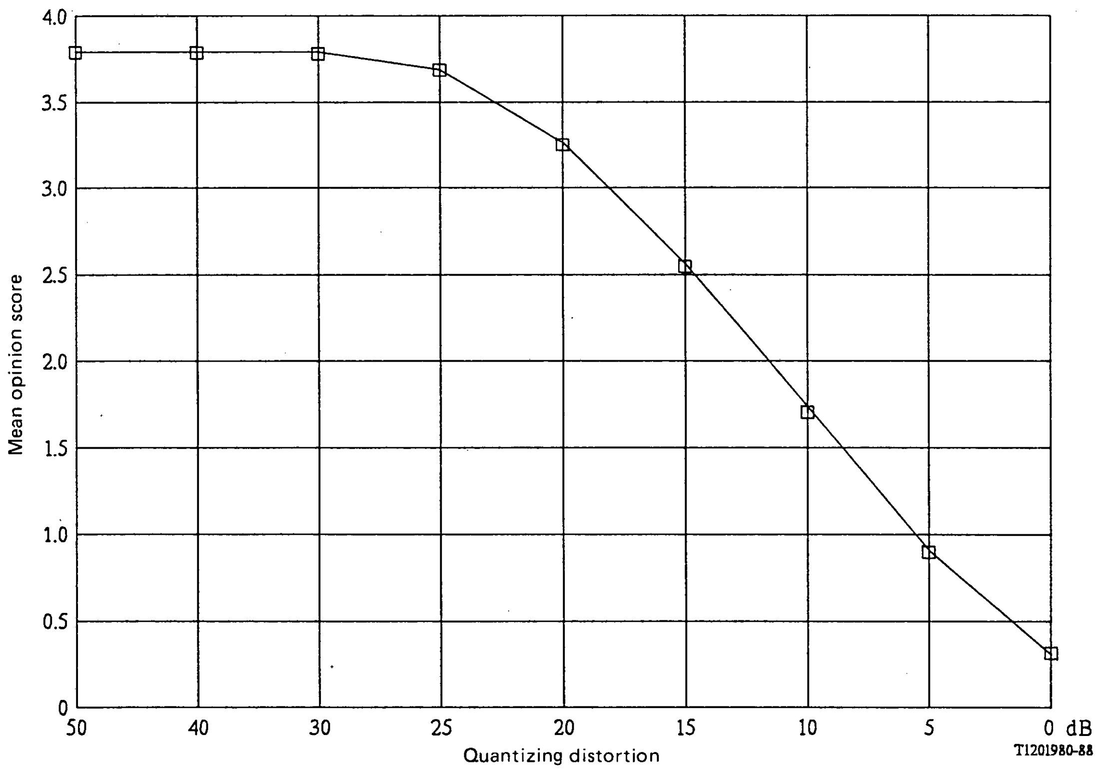


FIGURE A-4/P.11

Mean opinion score as a function of quantizing distortion

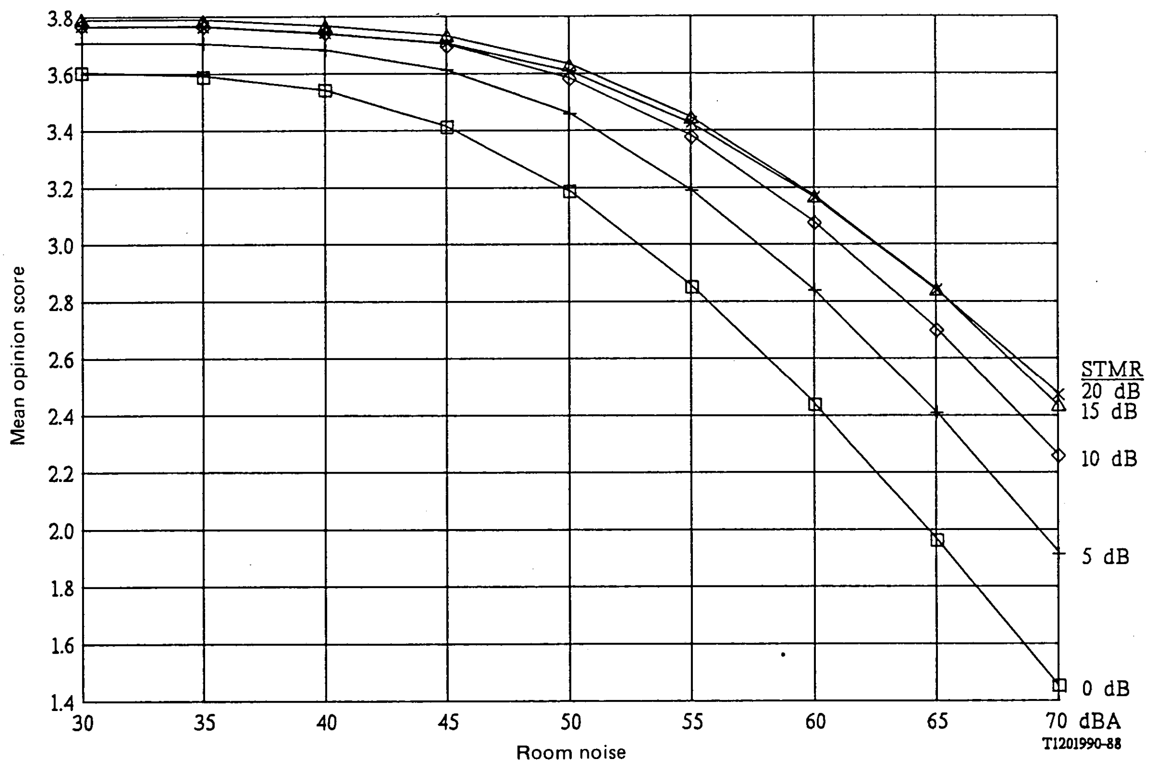


FIGURE A-5/P.11

Mean opinion score as a function of room noise and sidetone with a carbon microphone

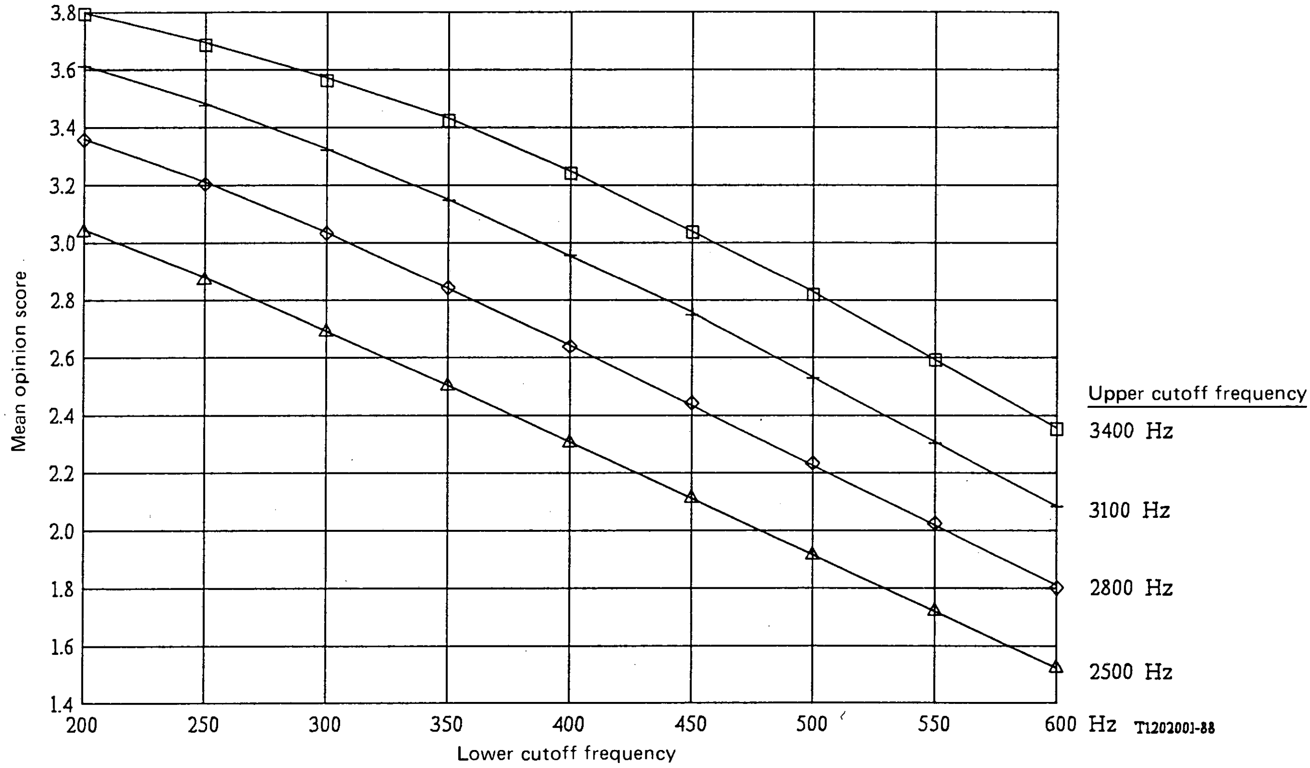


FIGURE A-6/P.11

Mean opinion score as a function of lower and upper cutoff frequencies

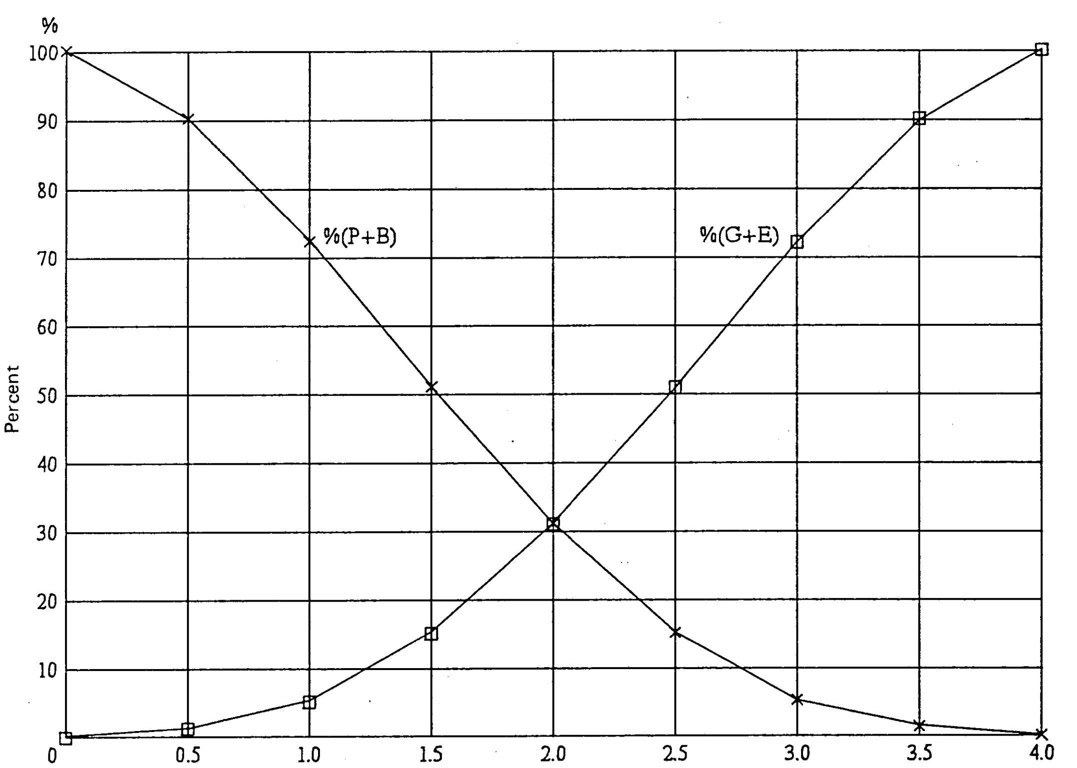


FIGURE A-7/P.11

Opinion relationships for the transmission quality index:
percent "G + E" and percent "P + B" as a function of mean opinion score

ANNEX B

(to Recommendation P.11)

Effects of attenuation distortion on transmission performance

B.1 *Effect of attenuation distortion on loudness and articulation*

The effect of attenuation distortion on loudness is more marked at a lower frequency band than at a higher one.

The effect of attenuation distortion on sound articulation is, contrary to loudness, more marked at a higher frequency band than at a lower one. Attenuation distortion equivalent values (I_L) and articulation equivalent loss values (I_A) are equivalent loss difference values referred to a system without frequency band restriction.

For both attenuation distortion equivalent and articulation equivalent loss values due to bandpass characteristics, it can be assumed that an additivity law of impairment values due to highpass and lowpass characteristics holds true, if each attenuation slope is steeper than 15 dB/octave.

These phenomena are induced based on the calculation and subjective test study results as shown in Figures B-1/P.11, B-2/P.11, B-3/P.11 and B-4/P.11.

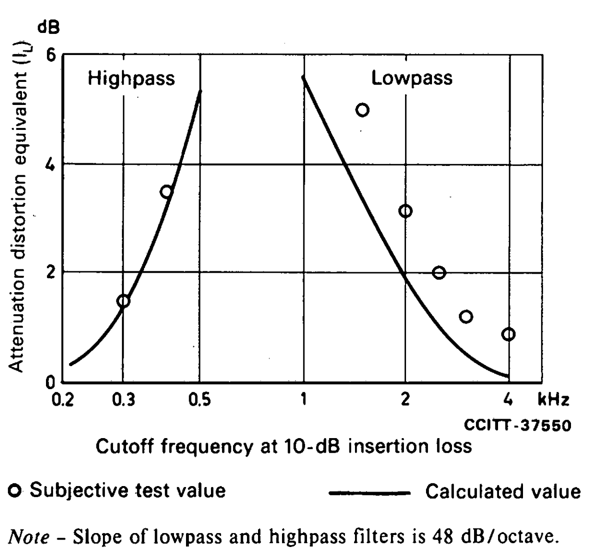
Note – Attenuation distortion equivalent and articulation equivalent loss described here are determined in reference to a complete telephone speech path without attenuation distortion junction.

B.2 *Effect of attenuation distortion on listening and conversation opinion scores*

The effect of attenuation distortion on listening and conversation opinion scores increases noticeably as the overall loudness loss of a connection decreases. This tendency can be more marked when circuit noise exists.

The effect of attenuation distortion on opinion scores is somewhat less than that of loudness loss, which is always dominant at any, particularly high overall loudness loss. However, its effect seems to be comparable to, or even larger than, that of noise under certain conditions, especially in connections of lower overall loudness loss.

See Figures B-5/P.11, B-6/P.11, B-7/P.11 and Table B-1/P.11.



Note – Slope of lowpass and highpass filters is 48 dB/octave.

FIGURE B-1/P.11

Cutoff frequency effect on loudness

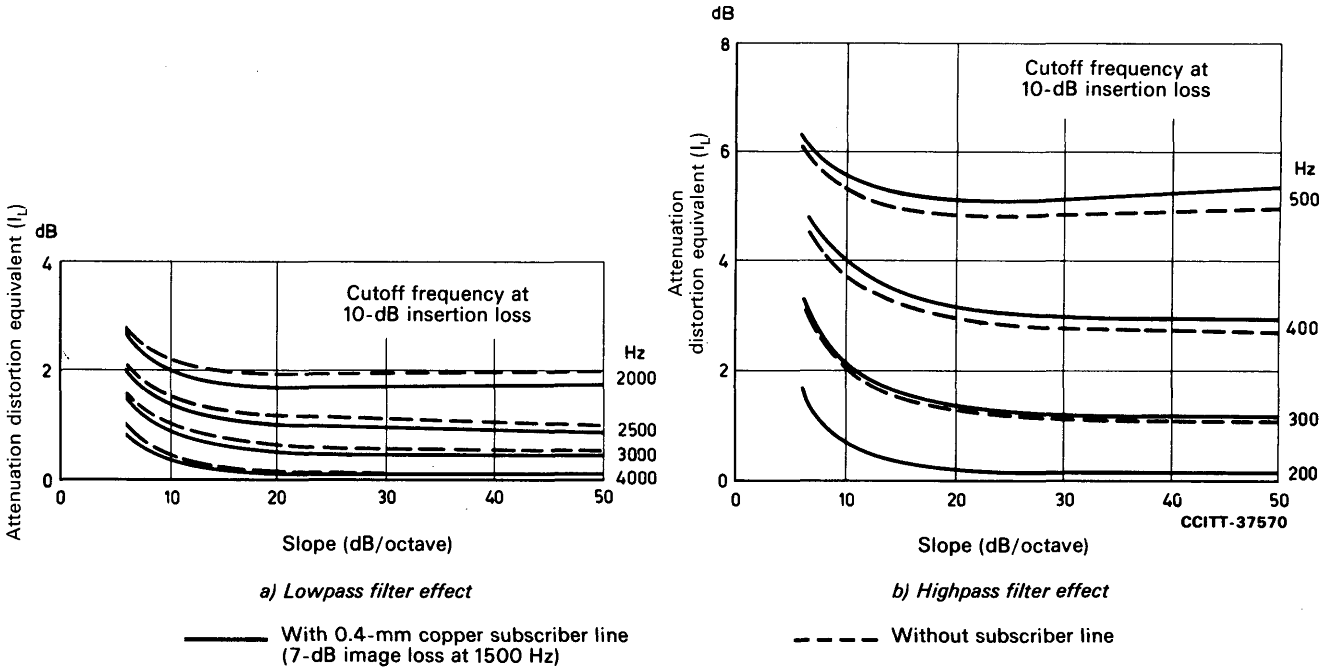


FIGURE B-2/P.11

Lowpass and highpass filter slope effect on loudness

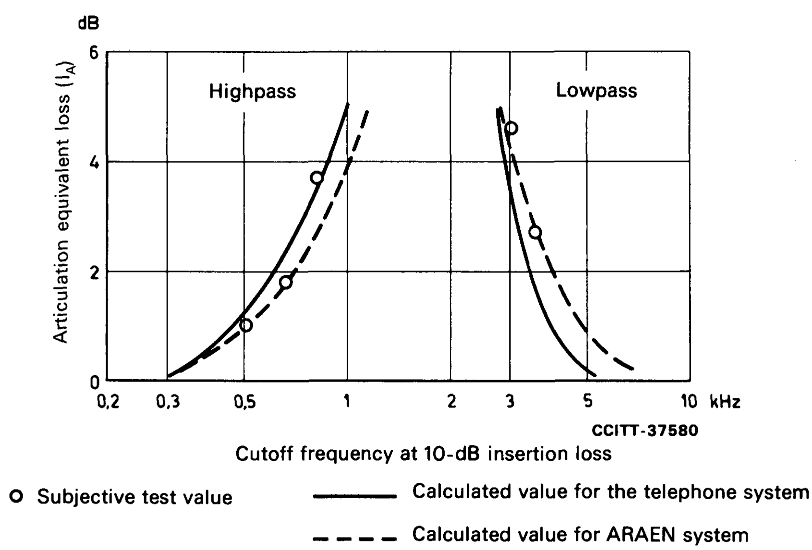


FIGURE B-3/P.11

Cutoff frequency effect on articulation

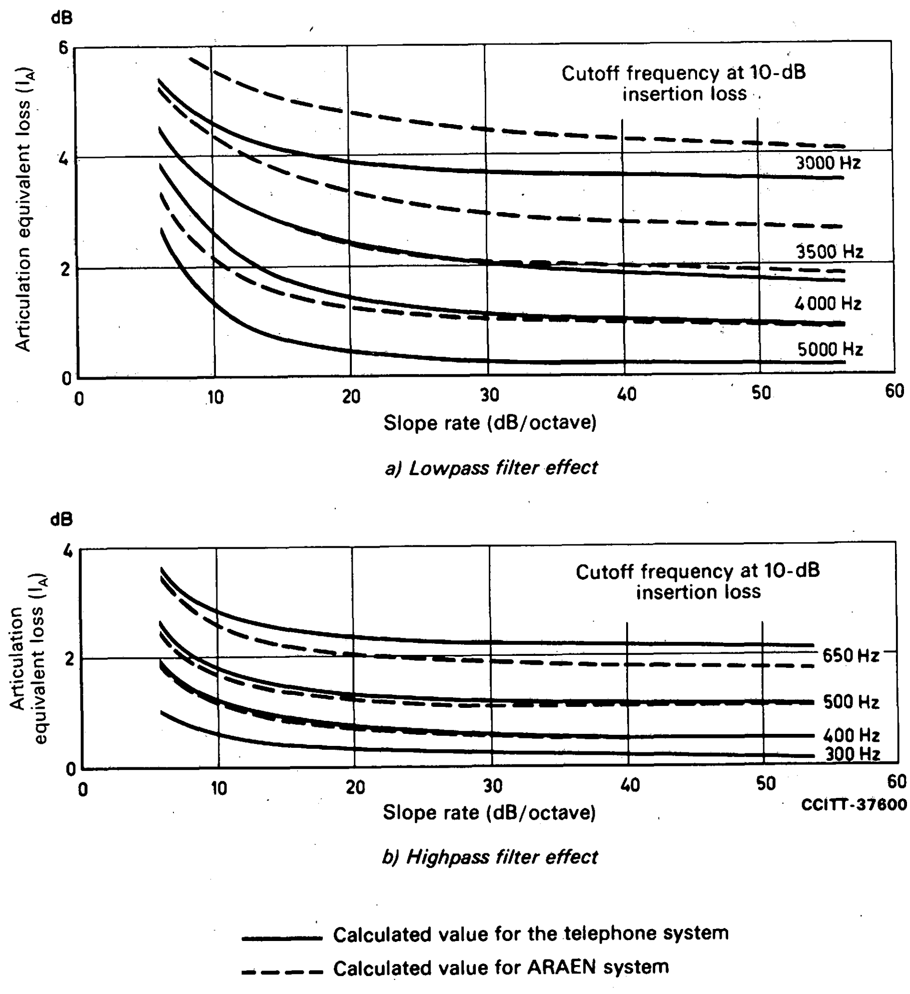


FIGURE B-4/P.11

Lowpass and highpass filter slope effect on articulation

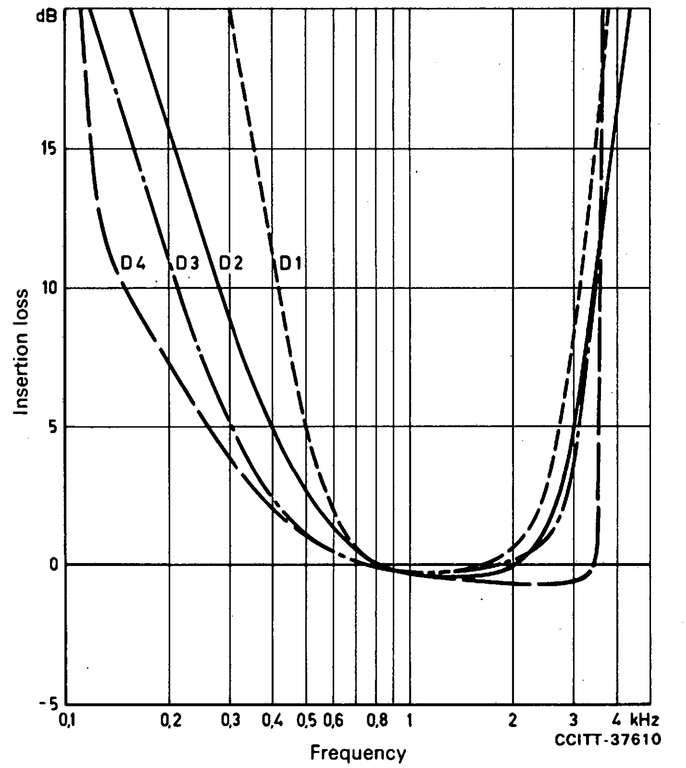
TABLE B-1/P.11

Opinion test conditions

No.	Item	Conditions of conservation opinion test using local telephone systems	Note
1	Junction loss	3, 13, 23, 29 dB	Measured at 800 Hz
2	Circuit noise level	ICN_0 [a] = −48.5 dBmp (14 000 pWp) −54.5 dBmp (3500 pWp) −60.5 dBmp (900 pWp) −78.5 dBmp (14 pWp)	Including exchange noise: −8 dB/octave spectrum characteristics
3	Room noise	50 dBA	
4	Sending and receiving end	Local telephone systems Telephone: Model 600 Subscriber line: 0.4 mm ∅, 7 dB at 1500 Hz Feeding bridge: XB exchange (220 + 220 Ω) Junction impedance: 600 Ω	SCRE + RCRE = 9.3 dB [b]
5	Attenuation distorsion	D1, D2, D3, D4 (Figure B-5/P.11)	

[a] Injected circuit noise referred to the input of a telephone receiving end with 0 dB receive corrected reference equivalent.

[b] SCRE = sending corrected reference equivalent, RCRE = receiving corrected reference equivalent.



D 1 12 4-wire circuits chain 95% limit characteristics, based on Figure 1/G.232, Graph No. 2B

D 2 12 4-wire circuits chain characteristics, based on Figure 1/G.132

D 3 Average characteristics of D4 and D2

D 4 SRAEN filter (Recommendations G.111 and P.11)

FIGURE B-5/P.11

Junction attenuation distortion characteristics for test conditions

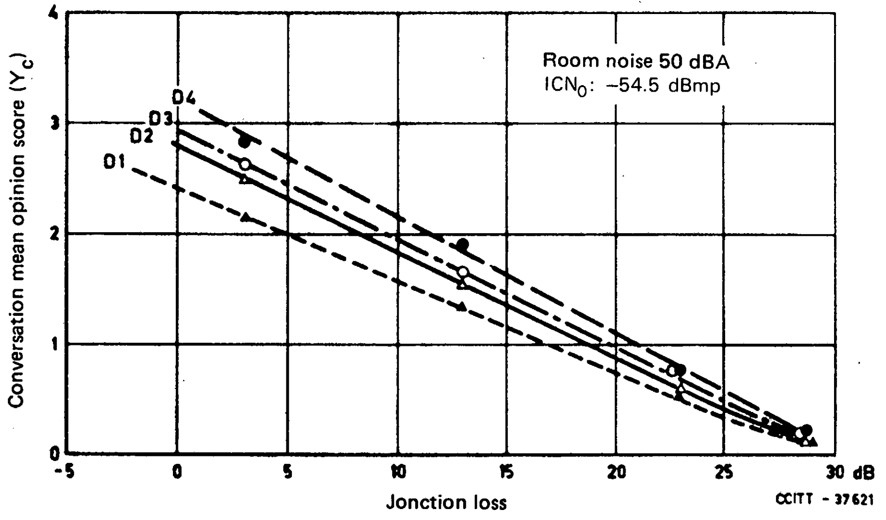


FIGURE B-6/P.11

Attenuation distortion effect on conversation opinion score

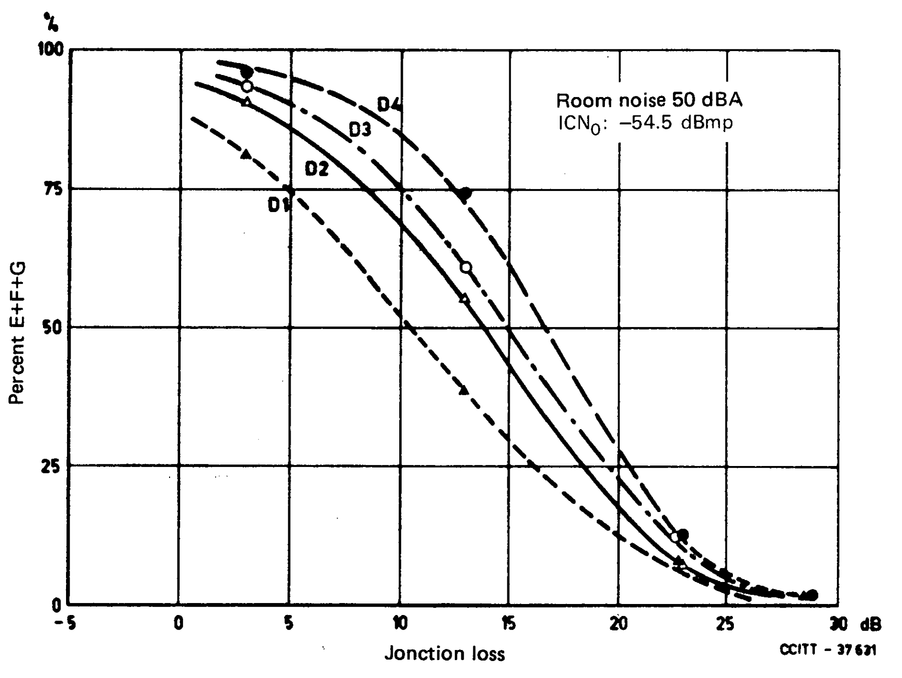


FIGURE B-7/P.11

Attenuation distortion effect on percent F, G and E in conversation test

B.3 *Examples of attenuation distortion characteristics effect*

TABLE B-2/P.11

Example of various methods to express attenuation distortion characteristics

Attenuation distortion	Characteristic parameters						Equivalent loss (dB)				
	Cutoff frequency (Hz)		Slope (dB/oct)		Insertion loss (dB)		Aspect 1		Aspect 2	Aspect 3	
	f_{L10}	f_{H10}	f_{L10}	f_{H10}	at 300 Hz	at 3.4 kHz	I_L	I_A	$I_{2.5}$	I_{Y_C}	$I_{\%FGE}$
D4	150	3500	7.0	300	3.8	0	0	0	0	0	0
D3	210	3400	10.0	31.5	5.2	10	0.8	0.3	–	2.3	1.8
D2	280	3300	10.7	29.1	8.8	10	1.2	0.5	1.8	3.8	2.8
D1	420	3100	22.2	31.1	20.0	15	3.2	2.2	4.2	7.8	6.3

I_L Attenuation distortion equivalent (calculated value).

I_A Articulation equivalent loss difference at 80% sound articulation (calculated value).

$I_{2.5}$ MOS equivalent loss difference at Y_{LE} = 2.5.

I_{Y_C} MOS equivalent loss difference at Y_C = 2.5.

$I_{\%FGE}$ Accumulated rating equivalent loss difference at 50% F, G and E.

B.4 *Evaluation method using the attenuation distortion unit (adu)*

The attenuation distortion unit (adu) may be used for evaluation of the attenuation distortion effect. However, a planning rule based on using an adu is not required.

Note – The attenuation distortion of a digital system is controlled by the existing planning rule based on using a quantizing distortion unit (qdu) because the methods used to assign qdu's to a digital system account for the effect of attenuation distortion. Therefore, there is no need for a planning rule based on using an adu.

The definition of attenuation distortion for one adu is shown in Table B-3/P.11.

TABLE B-3/P.11

Definition of attenuation distortion for one adu

Frequency (Hz)	Loss (dB)
200	1.57
300	0.40
400	0.12
500	0.08
600	0.06
800	0.01
1000	0
2000	−0.02
2400	0.05
2800	0.14
3000	0.17
3400	1.04

Note – This characteristic for one adu is based on Table A-4/G.113.

Sensitivity/frequency characteristics of local telephone systems (LTS) used to determine the effects of using adu's on speech quality are shown in Table B-4/P.11. These are intermediate reference system (IRS) characteristics without SRAEN filter characteristics. The IRS for each sending and receiving portion should be used as the sending and receiving portions of the network. For an ordinary telephone set, the differences in sensitivity/frequency characteristics are calculated from the IRS characteristics without SRAEN filter characteristics and transformed to adu numbers by the adu number rating method.

A rating method for attenuation distortion characteristics with regard to the number of adu's is described by the following equation:

$$N = \frac{1}{4}\left(\frac{A'_{300}}{A_{300}} + \frac{A'_{400}}{A_{400}} + \frac{A'_{500}}{A_{500}} + \frac{A'_{3000}}{A_{3000}}\right)$$

where:

N is the number of adu's

A'_f is the attenuation distortion of characteristics to be rated at frequency f (dB)

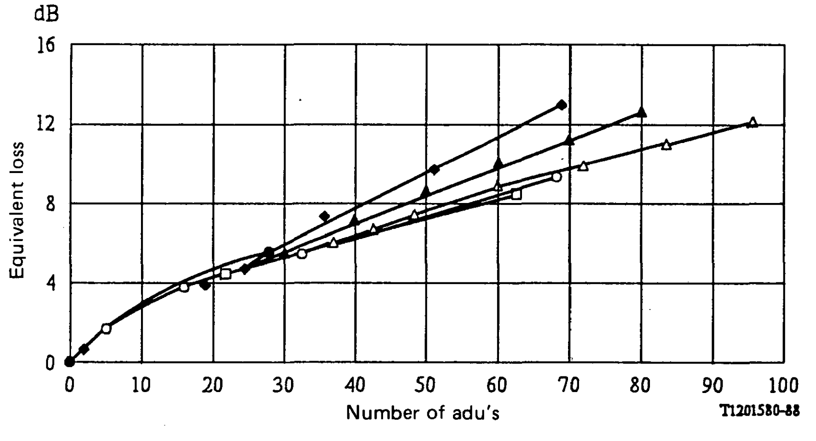
A_f is the attenuation distortion of one adu at frequency f (dB).

Opinion equivalent loss values for various numbers of adu's are shown in Figure B-8/P.11. Using the frequency characteristics shown in Tables B-3/P.11 and B-4/P.11, the reference point and number of adu's is calculated by the adu number rating method. According to Figure B-8/P.11, the total equivalent loss is approximately 0.15 dB per adu and is proportional to the number of adu's.

TABLE B-4/P.11

LTS sensitivity/frequency characteristic used to determine the effects of using adu's

Frequency (Hz)	Relative response (dB)	
	Sending	Receiving
100	−22.0	−21.0
125	−18.0	−17.0
160	−14.0	−13.0
200	−10.0	−9.0
250	−6.8	−5.7
315	−4.6	−2.9
400	−3.3	−1.3
500	−2.6	−0.6
630	−2.2	−0.1
800	−1.2	0
1000	0	0
1250	1.2	0.2
1600	2.8	0.4
2000	3.2	0.4
2500	4.0	−0.3
3150	4.3	−0.5
4000	0	−11.0
5000	−6.0	−23.0
6300	−12.0	−35.0
8000	−18.0	−53.0



Source

- o Annex A of this Recommendation
- □ China [1]
- Δ ATT [2]
- ● NTT [3]
- ◆ NTT [4]
- ▲ Total loss

Number of adu's	0	10	20	30	40	50	60	70	80
Junction loudness rating	0	1.0	1.8	2.4	3.1	3.7	4.3	5.1	5.9
Total loss	0	3.1	4.8	6.1	7.5	8.9	10.2	11.3	12.7

FIGURE B-8/P.11

Opinion equivalent loss value for various numbers of adu's

ANNEX C

(to Recommendation P.11)

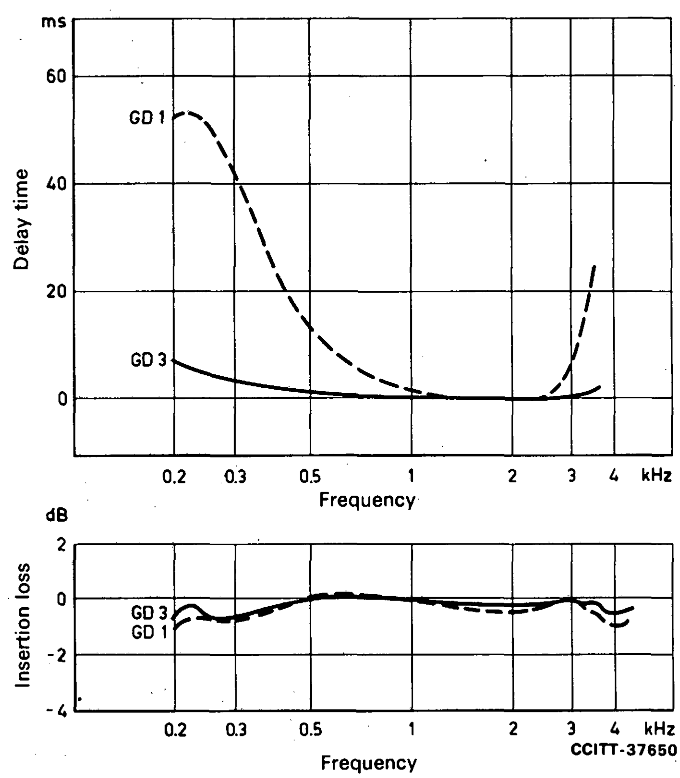
Effects of group-delay distortion on transmission performance

The effect of group-delay distortion is described as "ringing" at the upper part of a transmitted frequency band and as "speech blurred" at the lower part.

Absence of noise or attenuation distortion has such an influence as to hold the effect conspicuous throughout the possible overall loudness range of a connection.

However, its practical effect in a 4-wire circuit chain does not seem serious, since it is usually accompanied by closely related attenuation distortion.

See Figures C-1/P.11, C-2/P.11 and C-3/P.11.



GD1 : Approximated to 12-circuit chain 95% values
GD3: Approximated to typical modern one circuit value

Note – The test conditions are the same as those for the attenuation-distortion opinion test. The circuits modelling junction group-delay distortions used in the test are free from attenuation distortion.

FIGURE C-1/P.11

Junction group-delay distortion of test connection

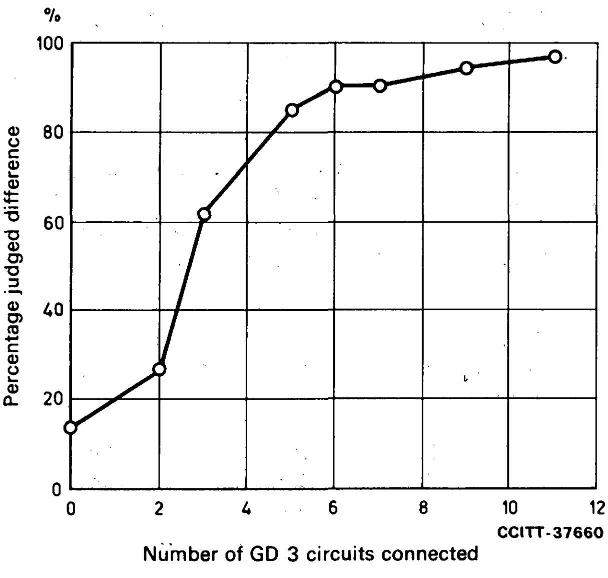


FIGURE C-2/P.11

Group-delay distortion detectability

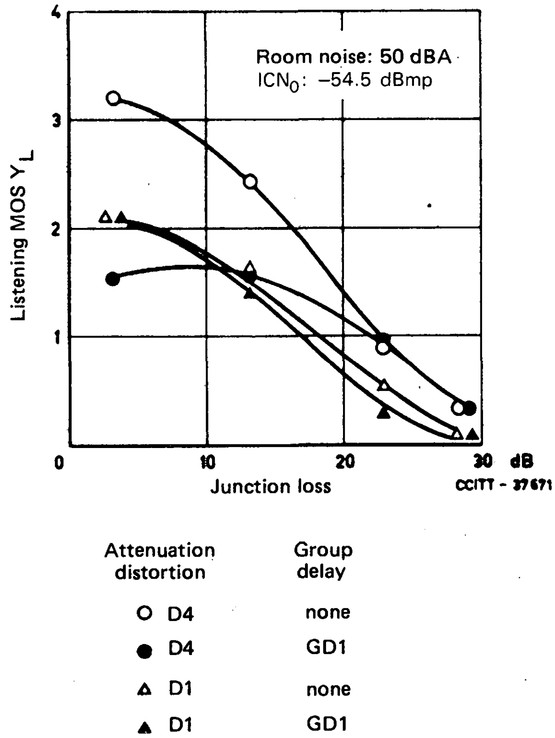


FIGURE C-3/P.11

Effect of group-delay distortion on listening opinion score

ANNEX D

(to Recommendation P.11)

Effects of carbon and linear microphones on transmission performance

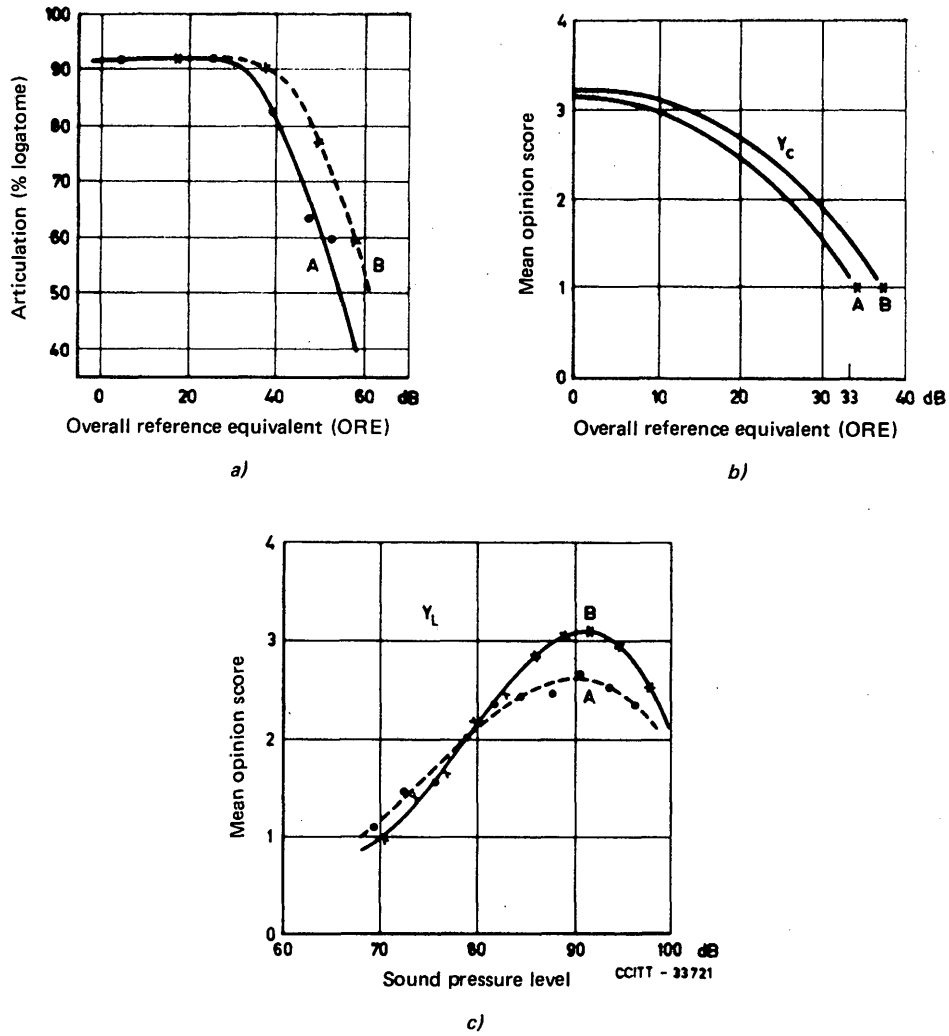
Information on the performance of carbon microphones as opposed to linear (non-carbon) microphones has been collected. The performance depends not only on differences in the content of non-linear distortion due to harmonics and intermodulation products but also on differences in amplitude/frequency distortion ("linear distortion") and amplitude/amplitude distortion (level-dependent sensitivity) between the two types of microphones.

Typical examples of results from comparative tests are given in Figure D-1/P.11. The diagrams show transmission performance measured as articulation or mean opinion score (for conversation or listening only) as functions of reference equivalent or speech level.

No general conclusion can be drawn from such results coming from different sources and dealing with various makes of microphones, because the individual effects of non-linear distortion and of frequency and amplitude-dependent sensitivity cannot be separated. Nevertheless, all three examples indicate some improvement of the transmission performance when a carbon-type microphone is replaced by a linear microphone.

In the particular example *c)* there is a significant improvement at optimum listening level while there is no difference (or even negative difference) at low listening levels. In that case, with room noise present and insufficient sidetone loss (sidetone reference equivalent 1-4 dB for this test condition) the inferior sensitivity of the specific type of carbon microphone to sound in the acoustic far-field may be an advantage.

For transmission over a bandwidth larger than the conventional telephone band – and in particular for loudspeaker listening – it is likely that there is a more noticeable improvement in sound quality if linear microphones are used instead of carbon microphones.



A Carbon microphone
B Linear microphone

Note – Frequency band: 300-3400 Hz, 50 dB(A) room noise.

FIGURE D-1/P.11

ANNEX E

(to Recommendation P.11)

Quantizing distortion of digital systems

To enable network planning for telephone speech transmission, it is convenient to assign appropriate weights to any nonstandard analogue/digital conversion process, transmultiplex pairs and processes introducing digital loss. An appropriate method is to consider that 1 unit of impairment is assigned to an 8-bit A- or μ-law codec pair to cover quantizing distortion. A planning rule provisionally agreed is to allow 14 units of impairment for an overall international connection, with up to 5 units for each of the national extensions and 4 units for the international chain. Such a rule would allow 14 tandem unintegrated 8-bit processes.

A subjective opinion model (see Supplement No. 3 at the end of this Volume) provides results which indicate that the Q[2)] for an overall connection with 14 unintegrated 8-bit systems in tandem is about 20 dB. The same model shows that one 7-bit system has the same Q as about three 8-bit systems. (This is based on the finding that subjective Q values for digital systems combine on a 15 $\log_{10}$ basis, i.e. 2 digital systems each with a Q = 24.5 dB would yield a Q = 20 dB when connected asynchronously in tandem.) It is recommended that until further information is available, 3 units of impairment (3 qdu) be assigned to a 7-bit system on speech transmission quality.

The provisional values given in Table E-1/P.11 for impairment unit assignment are recommended for planning purposes. These assignments are based on telephone speech considerations.

Note – These preliminary conclusions are based on a limited amount of information and the weights may be revised if new information becomes available.

TABLE E-1/P.11

Impairment unit assignments for telephone speech transmission

Process	Number of impairment units	Remarks
One 8-bit A-law or μ-law PCM	1	(Note 1)
7-bit PCM codec-pair (A-law or μ-law)	3	(Note 1)
One digital pad realized by manipulating 8-bit PCM code words	1	(Note 2)
One 32 kbit/s ADPCM-V	3.5	(Note 3)

Note 1 – For general planning purposes, half the values indicated may be assigned to either of the send or receive parts.

Note 2 – The impairment indicated is about the same for all digital pad values in the range 1-8 dB. One exception is the 6 dB A-law pad which introduces negligible impairment for signals down to about –30 dBm0 and thus attracts 0 units of quantizing distortion.

Note 3 – ADPCM-V = ADPCM with adaptive predictor (Recommendation G.721).

[2)] Q is the ratio of speech power to speech-correlated noise power determined subjectively by using the MNRU (Modulated Noise Reference Unit) (see the Recommendation P.81). Methods used for subjective assessment of codecs using the MNRU are outlined in Supplement No. 14.

ANNEX F

(to Recommendation P.11)

Effects of nonlinear distortion on transmission performance

The subjective effects of nonlinear distortion on real speech are highly dependent on the exact form of the nonlinearity. Figure F-1/P.11 gives some guidance on the degration introduced in terms of mean opinion scores obtained in actual subjective tests carried out by BNR in 1982 and 1986 and by NTT in 1986, for two forms of generalized nonlinearity namely, quadratic and cubic.

The main point to note is that, for a given amount of distortion (expressed in terms of the percentage of harmonic distortion of a sinusoidal signal having the same r.m.s. level as speech), the subjective effect of cubic nonlinearity is considerably more severe than that of quadratic nonlinearity.

The information given in Figure F-1/P.11 was derived from experiments based on a talker-to-listener path, and does not necessarily apply to nonlinear distortion occurring in a talker sidetone path, where there will be a masking effect of the undistorted speech signal.

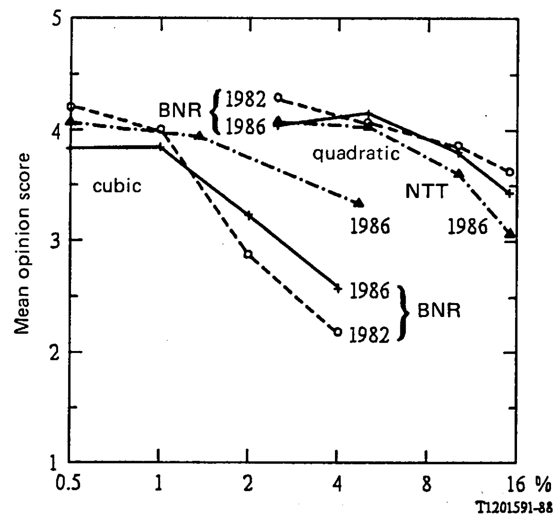


FIGURE F-1/P.11

Subjective ratings for nonlinear distortion

References

[1] CCITT – Contribution COM XII-No. 46, Study Period 1981-1984.

[2] CCITT – Contribution COM XII-No. 84, Study Period 1981-1984.

[3] CCITT – Contribution COM XII-No. 88, Study Period 1981-1984.

[4] CCITT – Contribution COM XII-No. 173, Study Period 1981-1984.

SUBJECTIVE EFFECTS OF DIRECT CROSSTALK; TRESHOLDS OF AUDIBILITY AND INTELLIGIBILITY

(Geneva, 1972; amended at Geneva, 1976, 1980; Malaga-Torremolinos, 1984; Melbourne, 1988)

1 Factors which affect the crosstalk threshold

The degree of audibility and intelligibility of a crosstalk signal depends on a large number of factors.

The main factors influencing the intelligibility of the vocal crosstalk signal are listed below.

1.1 *Quality of transmission of telephone apparatus*

The sending and receiving loudness ratings arre decisive factors. The same is true of the sidetone rating when room noise is present. The use of modern telephone apparatus with smooth frequency response curves is assumed.

1.2 *Circuit noise*

The circuit noise on the connection of the disturbed call must be taken into account. The noise level is measured by a psophometer equipped with a weighting network for telephone circuits, as described in Recommendation O.41.

1.3 *Room noise*

Room noise affects the ear directly through earcap leakage between the ear and the receiver and indirectly by sidetone. Sidetone also depends on operating conditions. Unlike circuit noise, the effect of room noise can be reduced to some extent by the user of the telephone. For this reason, and to allow for unfavourable cases, measurements on the audibility of crosstalk have been made with slight room noise as well as with negligible room noise. Because the audibility threshold is very sensitive to masking effects, «negligible» room noise means a noise level well below 10 dBA. The relatively low noise level of 40 dBA has a very marked masking effect and may therefore serve as an example of "slight" room noise.

1.4 *Telephone set noise*

In addition to the masking effects on crosstalk by circuit noise and room noise, the internal noise of the telephone set in the disturbed connection has to be considered. In modern telephone sets this noise is generated in the electronic circuitry (amplifiers, etc.) while in older sets the origin is noise from the carbon microphone. The internal noise can be expressed and treated as an equivalent circuit noise.

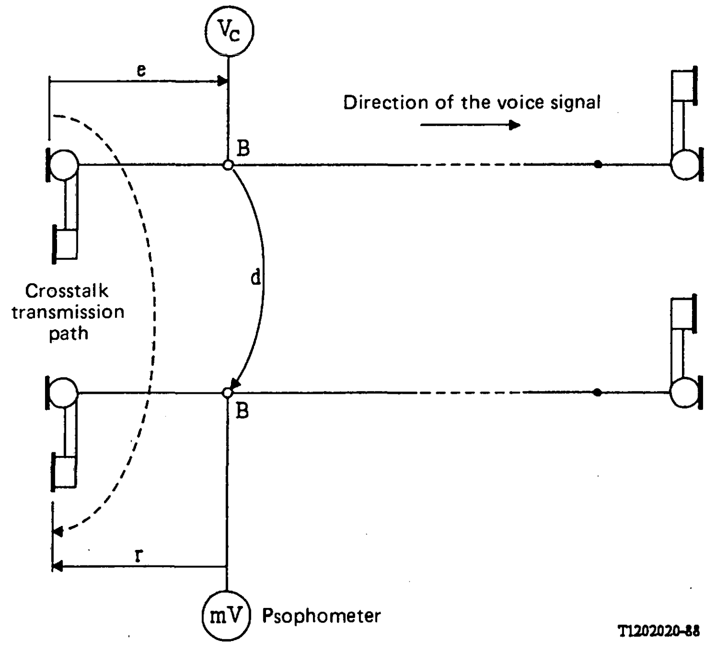
1.5 *Conversation on the disturbed connection*

During active speech on the disturbed connection, practical levels of crosstalk are inaudible. However, before the conversation starts or during long pauses in the conversation, it is possible for crosstalk to be heard and perhaps understood. In general, it would be unwise to plan on the basis that the disturbed connection is always active; accordingly, the information given in this Recommendation assumes no conversation on the disturbed connection.

1.6 *Crosstalk coupling*

The intelligibility of a vocal crosstalk signal depends largely on the nature of the crosstalk coupling, which is generally a function of frequency.

The loudness rating of the crosstalk transmission path – from the speech signal present on the disturbing line to the subscriber's set subject to the disturbance – can be divided into the loudness loss of the crosstalk path from the disturbing to the disturbed line and the receive loudness rating of the disturbed subscriber's set. Figure 1/P.16 illustrates this subdivision.



e sending LR of the disturbing subscriber's set
r receiving LR of the disturbed subscriber's set
d crosstalk path loudness loss
B terminals of subscriber's set
V_c conversational speech level

Disturbing and disturbed subscriber's sets at the same end: near-end crosstalk.
Disturbing and disturbed subscriber's sets at opposite ends: far-end crosstalk.

FIGURE 1/P.16

Subdivision of crosstalk transmission path

For a given speech level V_c, the intelligibility of the crosstalk signal depends on the loudness rating $d + r$. In Recommendation G.111, § A.4.4.4, the crosstalk receive loudness rating is defined as:

$$XRLR = RLR(set) + L_x$$

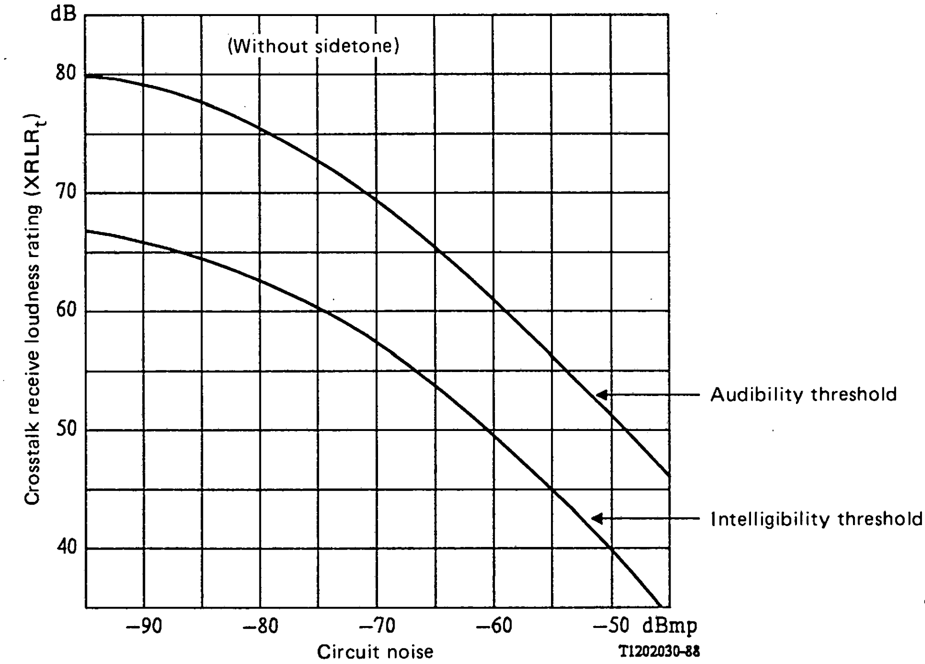
where *RLR*(*set*) refers to the disturbed telephone set.

The crosstalk loudness L_x is computed as a loudness rating but with the exponent m = 1, which is valid near the audibility threshold.

In the absence of further information, the value of L_x may be approximately taken as the attenuation measured or calculated at a frequency of 1020 Hz.

2 Median listener threshold of the audibility and intelligibility of vocal crosstalk

The curves in Figure 2/P.16 represent the crosstalk receive loudness rating corresponding to the threshold of audibility and intelligibility ($XRLR_t$) as a function of circuit noise. For planning purposes, it is recommended that room noise be regarded as negligible, which represents the most unfavourable condition.



Note 1 – The circuit noise is referred to the terminals of the subscriber's set. Room noise is assumed to be negligible.

Note 2 – The vocal level on the disturbing line is assumed to be −18 dBV active speech level.

FIGURE 2/P.16

Threshold value of crosstalk receive loudness rating as a function of circuit noise

The criterion for the threshold of audibility is that the presence of a speech signal is only just detectable but that no part of the speech can be understood. The criterion for the threshold of intelligibility is that single words or phrases can sometimes be understood while listening to a conversation.

The threshold curves represent median values for the two criteria such that in each case 50% of subscriber's opinions are respectively above and below the particular curve. The standard deviation for listeners has been observed to lie in the range 4 to 6 dB and a value of 5 dB is recommended for planning purposes. Typical response curves for a large sample of listeners for the threshold criteria are shown in Figure 3/P.16 (no circuit noise). The difference in XRLR between the two curves is about 12 dB.

The results of the original experiments (from which the curves in Figure 2/P.16 were drawn) were expressed in terms of speech level (e.g. in Volume Units (VU)) and on that basis showed a satisfactory degree of coherence.

However, earlier versions of Recommendation P.16 were based on the assumption that there is a fixed relationship between the sending loudness rating and the speech level on the line. This assumption required a correction in the range of 11 dB and is therefore not justified. Furthermore, speech levels expressed in Volume Units appear to differ systematically as measured in different countries on identical speech samples. Therefore, a fixed speech level on the disturbing line is assumed, independent of the send loudness rating (SLR) of that circuit.

The thresholds given in Figure 2/P.16 are based on the assumption that the speech level V_c under normal conversational conditions is −18 dBV active speech level (measured according to Recommendation P.56) at the terminal of the disturbing telephone set. This value is the estimated average of the conversational level in many countries at the send end of a connection with fairly high overall loudness rating [between the optimum and the maximum permitted (OLR)].

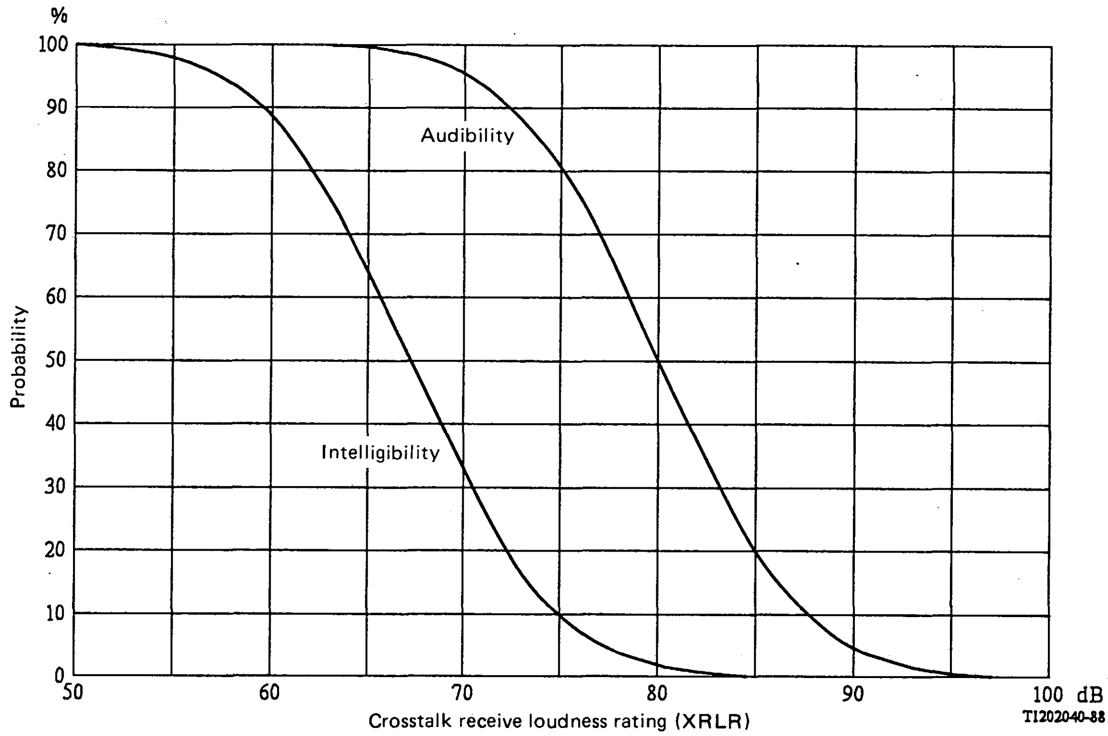


FIGURE 3/P.16

Listener distribution of crosstalk receive loudness rating required for different threshold definitions

The standard deviation of talking levels is fairly high. For calculation purposes a value of $\sigma = 5$ dB should be used.

To calculate the threshold value for a speech level different from -18 dBV, the $XRLR_t$ value should be corrected by the amount of the difference, with its sigh (higher levels require higher XRLR values, and vice versa).

The value $XRLR_t$ is the sum of the crosstalk path loudness loss and the receiving loudness loss on the disturbed line. In order to obtain the loudness loss of the crosstalk path, L_x, for a particular threshold value, the RLR(set) value has to be subtracted.

In general, for any speech level and receiving loudness rating, L_x is obtained from Figure 2/P.16 as:

$$L_x = XRLR_t - RLR(set) + (18 + V_c)$$

3 Effects of room noise

Room noise reaches the listener's ear both by leakage under the earcap of the telephone handset and by the sidetone path. For a given sidetone the room noise can be converted to an equivalent circuit noise by means of a transmission model such as described in Supplement No. 3. A family of conversion curves with sidetone loss as parameter is found in Figure 2 of this Supplement.

As an example, with a fairly high sidetone loss (the same as used in the previous version of Recommendation P.16) a level of 40 dBA room noise is equivalent to a circuit noise level of -85 dBmp. This noise level reduces the threshold XRLR value by about 8 dB. An additional reduction will in most cases be caused by earcap leakage.

However, the importance of this effect cannot be generally predicted, since it depends both on the shape of the earcap and on user habits.

4 Crosstalk probability

While the curves in Figure 2/P.16 present the median values for various noise conditions, the curves ub Figure 3/P.16 represent the probability of audible or intelligible crosstalk, in percent, for the negligible noise condition. Similar probability curves can be derived from the median values for any circuit noise condition by the use of cumulative normal distributions with a standard deviation of 6 dB.

In a more general case, the talker variance should also be added. The mean speech level used in the calculations may be chosen to be lower than the relatively high level assumed in Figure 2/P.16, e.g. -20 dBV, which is closer to the average level in the network. An example of such an overall probability calculation is given in Annex A.

The threshold values of crosstalk loudness rating given in this Recommendation can be used in different ways. One possible interpretation is to require all normal telephone connections (i.e. faulty connections excluded) to have crosstalk conditions between the two threshold criteria. This means that, on the one hand, there is no point in requiring a higher crosstalk attenuation than the one corresponding to the audibility threshold and, on the other hand, that the intelligibility threshold should not be exceeded.

Another interpretation is to set the requirement so that there is a given small probability (e.g. 5%) that intelligible crosstalk can be encountered with negligible room noise and with the lowest circuit noise level found in the network. In practice, noise conditions are more favourable in the sense that crosstalk quite often is masked by room and circuit noise to the extent of becoming inaudible. For the average of all connections the risk of intelligible crosstalk will therefore be much smaller than the given percentage for the most unfavourable condition.

Crosstalk requirements may not necessarily be the same for all parts of the network. Although the maintenance of telephone secrecy is primordial, the subscriber is more likely to make a severe judgment on crosstalk in a local call taking place in his immediate environment and in which indiscretion due to crosstalk may have unfortunate social consequences. The problem of "social crosstalk" is dealt with in [1].

In practice, simultaneity of speaking on the disturbing line and listening on the disturbed line (during conversation pauses) is not present in all cases. Information concerning this topic and showing how ti calculate the probabilities concerned will be found in [2].

As guidelines, the probabilities of subscribers encountering potentially intelligible crosstalk should not be worse (i.e. higher) than the following:

- own exchange calls: 1 in 1000,
- other calls: 1 in 100.

Note – The fundamentals of calculating crosstalk probability in general are considered in Recommendation G.105.

ANNEX A

(to Recommendation P.16)

Example of probability calculation

The probability of understanding single words of a conversation overheard by crosstalk may be calculated for a listener chosen at random from a population of subscribers. The result of such a calculation can be used as a basis for establishing rules for, inter alia, the minimum required crosstalk attenuation between subscriber lines in a national network.

In order to demonstrate the method of using the information given in this Recommendation to calculate the probability of encountering (intelligible) crosstalk, the following assumptions may be made:

Mean speech level $V_c = -20$ dBV;

Receive loudness rating of telephone sets $RLR(set) = -6$ dB;

No room or circuit noise;

Standard deviation of talking levels $\sigma_T = 5$ dB;

Standard deviation of listener response distribution $\sigma_L = 6$ dB;

Standard deviation of $RLR(set)$ $\sigma_s = 1$ dB.

The threshold value for crosstalk intelligibility without noise, taken from Figure 2/P.16 is $XRLR_t = 67$ dB.

According to the formula at the end of § 2 and with the given assumptions, the required median crosstalk path loudness loss becomes:

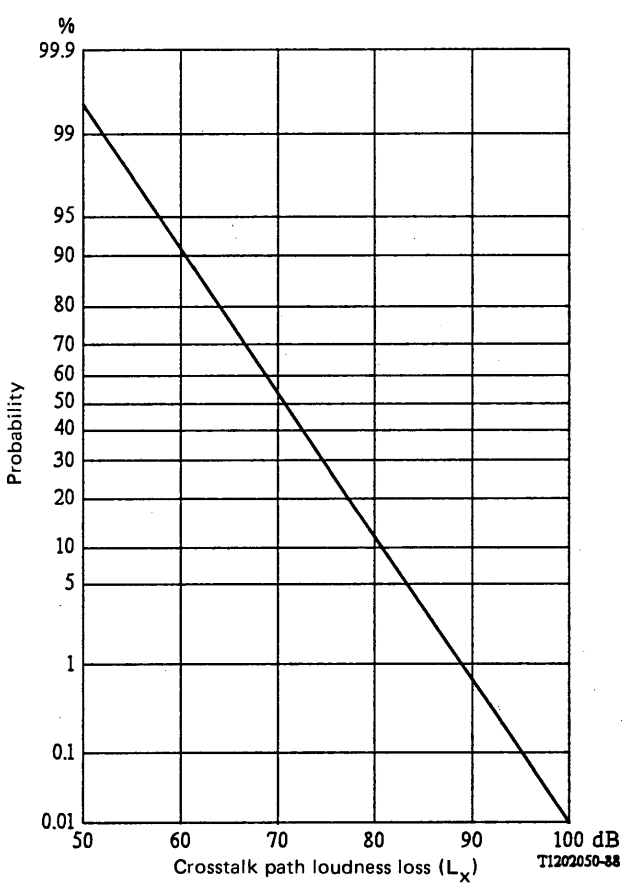
$$L_x = 67 + 6 - 2 = 71 \text{ dB}.$$

The total standard deviation of the probability function is:

$$\sqrt{\sigma_T^2 + \sigma_L^2 + \sigma_S^2} = 7.87 \text{ dB}$$

With these values of L_x and σ, a cumulative normal distribution function as in Figure A-1/P.16 can be drawn. The function indicates the probability that a listener can understand single words if crosstalk for a specific value of the crosstalk path loudness loss. For example, for $L_x = 75$ the probability is 30%. On the other hand, to obtain only 5% probability a crosstalk path loudness loss of 84 dB would be necessary. For 1% probability, 89 dBwould be required, as well as 95 dB for 0.1% probability.

This calculation was based on some typical values of speech level and receiving sensitivity under noise-free conditions. Similar calculations can easily be made with other data, also including the effects of noise. For a realistic estimation of the probability of intelligible crosstalk for subscribers in general, some statistical distribution of circuit noise (and possibly of room noise at the subscriber's locations) will have to be assumed.



V_C mean = –20 dBV; RLR (set) = –6 dB; standard deviation = 7.9 dB

FIGURE A-1/P.16

Probability of understanding single words of an overheard conversation as a function of the (weighted) crosstalk path loudness loss L_X

References

[1] WILLIAMS (H.), SILOCOCK (W. W.), SIBBALD (D.): Social crosstalk in the local area network, *El. Comm.*, Vol. 49, No. 4, London, 1974.

[2] LAPSA (P. M.): Calculation of multidisturber crosstalk probabilities, *BSTJ*, Vol. 55, No. 7, New York, 1976.

SECTION 2

SUBSCRIBERS' LINES AND SETS

Recommendation P.30

TRANSMISSION PERFORMANCE OF GROUP AUDIO TERMINALS (GATs)

(Melbourne, 1988)

1 Introduction

Group Audio Terminals (GATs) are terminals which have been specifically designed to be used by several users.

GATs cover a wide range of products ranging from the hands-free telephone when it is used by several users, to the sophisticated teleconference studio.

The CCITT recommends that GATs satisfy the specifications [1] in this Recommendation.

GATs must also comply with Recommendation P.34 as far as loudness is concerned, when they are connected to the telephone network. If they use voice-activated circuits, Recommendation P.34 may also be applied. Such terminals are sensitive to the acoustics of the location where they are utilized and they may resort to sophisticated acoustical echo processing devices.

The first generation of GATs will operate mainly on 4-wire digital networks and will make use of the wideband (WB) speech coding algorithm specified in Recommendation G.722. Such terminals urgently need specifications that can be based on the present Recommendation.

A typical GAT configuration is represented in Figure 1/P.30.

Such a terminal includes one or several microphones, one or several loudspeakers, sending and receiving amplification. Optionally, it includes a sound-managing and mixing device to the loudspeakers and from the microphones, a coder-decoder for digital networks, a voice-activated gain processing device and an echo processing device.

The location where the GAT is to be used is very important. Several measurements defined in this Recommendation have to be made at the location where the GAT is to be used. These are referred to as "in situ" measurements. They are to be made with the full complement of equipment in the conference room, but with no conferees present.

The present Recommendation is devided into three parts:

- interconnection specifications,
- transmit specifications,
- near-end specifications.

[1] The specifications in this Recommendation are subject to future enhancement and therefore should be regarded as provisional.

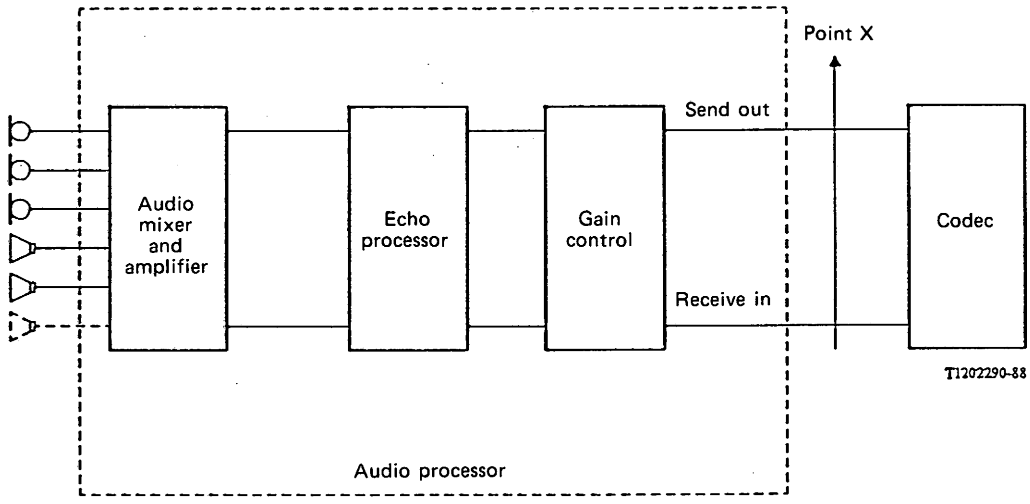


FIGURE 1/P.30

Group audio terminal configuration

Two test signals are used in this Recommendation:

- an acoustic test signal as defined in Recommendation P.50 (see Note): i.e. an artificial voice as defined in Recommendation P.50 produced by a sound source (an artificial mouth) as described in § 2 of Recommendation P.51 and,
- an electric test signal whose long-term spectrum is identical to the acoustic signal; when applied by a source with a matched internal resistive impedance, it provides a level of −22 dBV.

Both test signals are filtered in the transmission system bandwidth.

Note – The preferred acoustic signal to be used in the measurements for the audio alignment is defined in Recommendation P.50. However, other signals such as speech-shaped noise or pink noise may be used in some applications.

2 Interconnection specifications

These specifications are the basic requirements for a GAT to be connected to a network and to allow communication between several locations.

2.1 *Sending sensitivity*

2.1.1 *Wideband GATs*

For wideband applications, the transmission characteristics of the audio-channel shall be in accordance with Recommendation G.722.

2.1.1.1 *Send side alignment*

The sound source is positioned over the edge of the conference table on the centre line of each conferee's position, as defined in Recommendation P.34 (see Figure 3/P.34), and delivers a signal which complies with Recommendation P.64 [i.e. −4.7 dBPa at the mouth reference point (MRP)].

During the send side alignment the microphones of the GAT shall be positioned on the table as in real use.

The microphone gain controls must be adjusted to achieve, for each position of the source, an output line level of − 22 (± 2) dBV at point X (see Figure 1/P.30), assuming the signal recommended in Recommendation P.50 is used. This value takes account of an 18 dB peak factor of the speech signal and 6 dB for the variations between speakers and the variations due to conferees' movements.

2.1.2 *GATs connected to the public switched telephone network*

Such terminals must comply with Recommendation P.34.

2.2 *Stability test*

The GAT shall have a minimum stability margin of 3 dB when the microphone and loudspeaker paths are looped at reference point X in Figure 1/P.30 and the sound source is activated as described in § 2.1

During the measurement, the volume control shall be in maximum position.

3 Transmit quality specifications

These specifications limit the degradations induced on the network by a GAT.

3.1 *Electro-acoustical specifications*

3.1.1 *Microphone*

The electro-acoustical characteristics of the microhpones should conform to IEC Publication 581-5.

3.1.2 *Octave band measurements*

In situ measurement of the overall transmission frequency response characteristic is recommended. It is defined as the difference between the octave spectra of the electrical signal at the X interface and the acoustic excitation at the MRP. The artificial mouth is positioned as in § 2.1.1.

In order to prevent excessive fluctuations of the frequency response of the system, and since the measurements are performed on site, octave band measurements are recommended in the range 125 Hz to 4 kHz.

The sum of the absolute differences between the measured values and their average should be as low as possible. A practical target of 10 dB is achievable.

3.2 *Echo performance*

3.2.1 *Acoustic echo control*

To get satisfactory suppression of acoustic echoes it is necessary to provide the audio processor with either an echo canceller or an echo suppressor. The echo cancellation technology is recommended if highest possible speech quality performance is aimed at. However, it is recommended always to complement echo cancellation with a mild echo suppression, in order to prevent the undue transmission of room background noises when no talkers are active in the room. This condition should particularly be met in multi-conference environments.

3.2.2 *Echo return loss*

The echo return loss of the audio system shall be measured at reference point X of Figure 2/P.30, with the volume control in maximum position. When the electric test signal, as specified in § 1, is applied to the input port (receive in), the level measured at the output port (send out) shall not be higher than −62 dBV.

An acoustic echo loss of 40 dB includes a margin of 5 dB in order to provide an echo return loss of 35 dB when several GATs are used in a conference situation. This value of 35 dB should be understood as a minimum value. The long-term target value for the acoustic echo loss must be considered as being 45 dB (especially, to take into account the case where a handset is connected to a hands-free terminal). This value is known to prevent any subjective degradations due to delayed acoustic echo [1, 2]. The level measured at reference point X will then be −72 dBV.

Note – The echo canceller shall permit double-talk with negligible speech quality degradation (under study with Question 2/XII).

3.3 *Electrical noise*

The electrical noise emitted by the GAT at the reference point X should be less than −55 dBm, within the transmission bandwidth. No component outside the band should exceed 20 dB above the noise level in the band.

The measurement must be done with no conferees in the room and without incoming signals on the receiving side of the equipment in order not to activate the microphone circuits.

The noise emitted by the GAT at the reference point X when the microphones are active should be no more than −50 dBm. It must be measured by forcing the system into the emission mode as if one speaker were active in the room.

3.4 *Reverberated field picked up by the microphone*

For this measurement, the sound source is positioned in order that the distances between the sound source and all the microphones greater than three times the distance between the microphone and the position defined for the send side alignment. It is also recommended that the source be, at least, one meter from the walls. Then the signal measured at point X shall be not more than −29 dBV (this accounts for a direct-field over reverberated-field ratio of 6 dB [3]). It must be measured by forcing the system into the emission mode as if one speaker were active in the room. The test must be performed for each microphone in the room.

Basic requirements for the choice of the conference room, for its acoustical treatment and for the positioning of microphones and loudspeakers can be found in Supplement No. 16.

4 Near-end quality specifications

This part of the Recommendation tests the minimum specifications intended for the local users.

4.1 *Electro-acoustical specifications*

4.1.1 *Louspeakers*

The electro-acoustical characteristics of the loudspeakers should conform to IEC Publication 581-7.

4.1.2 *Octave band mesurements*

In-situ measurement of the overall reception frequency response characteristics is recommended. It is defined as the difference between the octave spectra of the acoustic signal delivered by the loudspeaker(s) at the listening positions and the input electric signal at the X interface.

The sum of the absolute differences between the measured values and their average should be as low as possible. A practical value of 12 dB is achievable.

4.2 *Receiving sensitivity*

4.2.1 *Volume control*

The audio conference terminal shall be provided with a volume control. The gain at maximum position should conform to § 4.2.2. The volume control should ideally be linked to the echo control mechanism.

4.2.2 *Receiving side alignment*

4.2.2.1 *Wideband GATs*

The electrical test signal is connected to the input port of the system. The receiving gain shall be adjusted in order to reach a sound pressure level of at least 65 dB and 20 dB above the acoustical noise level at the MRP. The alignment procedure should be performed with the volume control in the maximum position.

4.2.2.2 *GATs connected to the analogue public switched telephone network*

Such terminals must commply with Recommendation P.34.

References

[1] CCITT – Contribution COM XII-No. 170, Study Period 1985-1988

[2] CCITT – Contribution COM XII-No. 171, Study Period 1985-1988

[3] CCITT – Contribution COM XII-No. 172, Study Period 1985-1988

Recommendation P.31

TRANSMISSION CHARACTERISTICS FOR DIGITAL TELEPHONES

(Melbourne, 1988)

This Recommendation deals with sending and receiving loudness ratings, sidetone masking rating, listener sidetone rating, and sending and receiving sensitivity/frequency characteristics. Other important characteristics are still under study.

1 Sending loudness rating (SLR) and receiving loudness ratings (RLR)

In view of Recommendation G.111, § 3.2, the following values are recommended:

- as a short-term objective, nominal values of SLR in the range 5 to 11 dB and nominal values of RLR in the range -1 to 5 dB;
- as a long-term objective, the following nominal values for SLR, 8 dB and for RLR, 2 dB.

Note 1 – The recommended values for SLR and RLR do not imply that echo control in the network can always be avoided.

Note 2 – The acoustic loss in the telephone set is an important factor in the echo path and will need careful consideration. A volume control in the telephone set will decrease the echo loss by the same amount as the gain is raised.

Note 3 – For digital telephones connected to a digital PABX (to which analogue telephones may also be connected), values at the lower end of the ranges above might be needed. The reason is to give customers the same receiving level as they are used to having with the analogue telephones. A receiving volume control might be considered.

2 Sidetone masking rating (STMR) and listener sidetone rating (LSTR)

In view of the following considerations:

- the optimum STMR for conditions free from echo;
- the sidetone masking effect on talker echo at short delays;
- the difficulties of high ambient noise conditions;
- what subscribers are used to having with present analogue sets,

the following values are recommended [1]:

- nominal values of STMR in the range 10 to 15 dB;
- nominal values of LSTR >15 dB,

(No maximum values for LSTR need to be imposed.)

Note – These values may be modified when information becomes available on the effects of short delay echo (e.g. 10 ms).

3 Sending and receiving sensitivity frequency characteristics for digital telephones

In view of the following considerations:

- the compatibility with analogue telephones in a mixed analogue digital network;
- the absence of line-length-dependent frequency distortion to be compensated for as with analogue telephones;
- the aim to achieve the best possible overall quality with the digital telephone,

[1] The specifications given here are subject to future enhancement and therefore should be regarded as provisional

sending and receiving sensitivity/frequency characteristics as specified below are recommended:

- a substantially flat receiving frequency response S_{JE} between 300 Hz and 3 400 Hz should be chosen;
- a nominal sending frequency response S_{MJ} rising with a slope within the area indicated in Figure 1/P.31 should be striven for;
- below 200 Hz, the send slope should fall by at least 6 dB/octave.

Note 1 – S_{JE} and S_{MJ} are normally estimated from measurements of S_{Je} and S_{mJ} according to Recommendation P.66.

Note 2 – An expansion of the lower frequency range to 200 Hz will increase the naturalness of the speech.

Note 3 – The normal considerations for anti-aliasing filters must be applied to the frequency responses.

Note 4 – Marked peaks in the responses might cause stability problems and should therefore be avoided.

Note 5 – The preferred curves for S_{JE} and S_{MJ} defined in this way should be considered as a design objective. Individual microphone and receiver curves will, for several reasons, deviate more or less from the "ideal" curves. However, it is hardly possible to specify in a Recommendation concerning desirable frequency characteristics how much and in which way individual response curves may deviate from the target curve, without becoming unacceptable. For type approval of telephone sets, it is generally necessary to specify limits for the shape of sending and receiving frequency curves nationally, in the same way as tolerance limits for loudness ratings are usually specified. These limits are based on technical considerations as well as on cost of implementation, manufacturing tolerances and other economic factors.

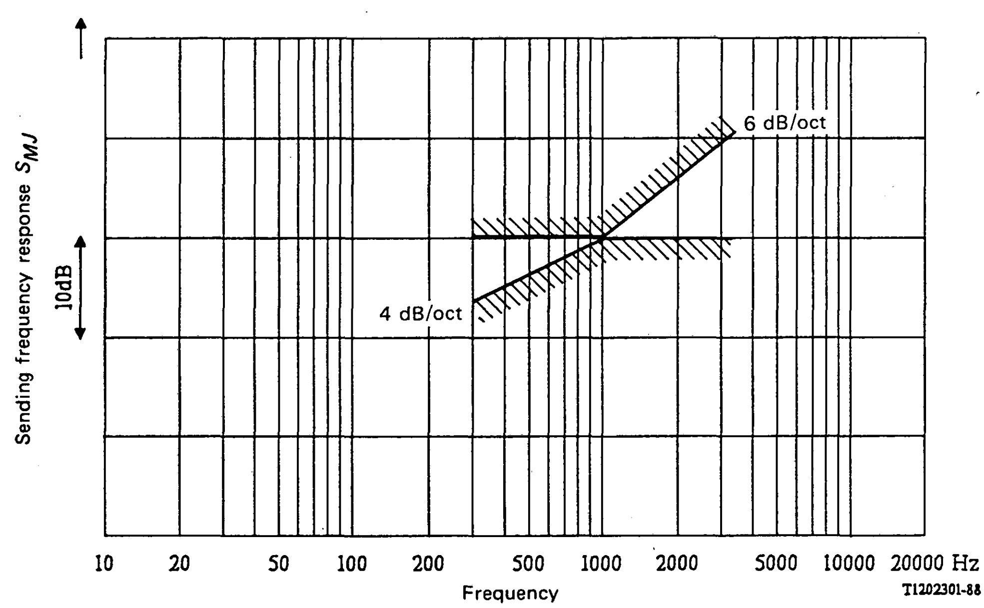


FIGURE 1/P.31

Recommendation P.32

EVALUATION OF THE EFFICIENCY OF TELEPHONE BOOTHS AND ACOUSTING HOODS

(Melbourne, 1988)

The purpose of this Recommendation is to define the methods of measurement to evaluate the efficiency of either acoustic hoods or telephone booths intended to improve the quality of telephone transmission in noisy environments. In addition to the improvement of the transmission quality during a conversation between two users, this Recommendation takes into consideration the need to guarantee speech privacy for the user speaking from the acoustic hood or the booth with respect to a listener situated on the outside of the telephone booth.

1 Evaluation methods

The efficiency of a telephone booth or an acoustic hood can be evaluated using either subjective or objective measurements.

The objective measurements suitable for that purpose are those based on the acoustic insulation (for example in a reverberation chamber) resulting from the difference between the sound levels registered inside and outside the telephone booth or vice-versa. As the acoustic characteristics vary outside and inside the telephone booth, the acoustic insulation obtained in each case (noise source outside or inside the telephone booth) is not the same. In addition, if we consider open telephone booths or acoustic hoods, the measurement of the acoustic insulation gives results not correlatable to the ones obtained by means of a subjective evaluation of the booth performance.

A subjective measurement of the efficiency of booths or acoustic hoods consists in determining the intelligibility index inside the booth, in conditions of external noise (room noise, road noise, etc.). This measurement can also be objectively obtained by calculating the articulation index, by using, for example, the Kryter method as indicated in Annex A.

Another method used to measure subjectively the efficiency of booths and acoustic hoods consists in evaluating the intelligibility threshold variation observed between the intelligibility inside and outside the booth placed in a noisy ambient.

The performance of booths and acoustic hoods, related to the user's privacy while speaking inside the booth, can be subjectively evaluated by measuring the intelligibility of the coversation from the inside to the outside of the booth or by using an objective measurement such as calculating the articulation index (according to Kryter's method for example) outside the booth under specific noise conditions.

Since the intelligibility inside the booth is also a function of the sidetone of the telephone set used, a simple measurement of acoustic insulation which does not take into consideration the intelligibility reduction caused by the sidetone cannot furnish correct evaluations on the improvement of transmission quality due to telephone booths, or acoustic hoods.

Bearing in mind the following observations:

1) international telephone communication can be originated from telephone sets installed in noisy ambients and protected by booths or acoustic hoods;
2) there are no measurement methods recommended for evaluating the transmission quality improvement resulting from the use of the telephone booth;
3) an evaluation of the booth efficiency, based only on the acoustic insulation obtained by traditional methods (acoustic attenuation of the panels of the booth) is not always correlated to the subjective evaluation of the booth performance;
4) subjective measurements either of the intelligibility or of the intelligibility threshold variation give the possibility of evaluating the efficienty of a booth, but are time-consuming and expensive and also require a qualified and well-trained operator team;
5) there are no recommendations giving criteria relating the employment of the booths to the ambient noise level, in order to determine an acceptable quality of transmission,

methods of measurement as specified below are recommended[1]:

a) evaluating the efficiency of telephone booths and acoustic hoods taking into consideration the intelligibility index, obtained from a listener inside the booth with the external ambient noise having a certain acoustic spectrum;

b) calculating the intelligibility index inside the telephone booth or the acoustic hood by means of the objective method defined in § 3, taking into consideration the acoustic attenuation of the booth and the sidetone of the telephone set used. This objective method allows a rapid evaluation of the booth performance, sufficiently precise for practical purposes;

c) considering the logatom intelligbility as an evaluation criterion related to the booth performance, calculated by means of the articulation index (AI). The conversion from AI to logatom intelligibility is language-dependent and it shall be performed with the appropriate relation;

d) evaluating the booth and the acoustic hood at the conditions of utilization, that is, when a user is speaking from the inside using a telephone set with a determined sidetone and with an external ambient noise having an average intensity level and a certain acoustic spectrum, both already known.

2 Definition and descriptions of parameters of calculation

Telephone conversations taking place in conditions of ambient noise are affected by ambient noise through three different paths:

1) acoustic noise (N_a) at the ear which is not engaged in the telephone call;

2) acoustic noise (N_b) at the ear which is engaged in the telephone call, determined by the acoustic leak between ear and handset;

3) noise picked up by the microphone and directed by sidetone (N_s) to the ear which is engaged in the conversation.

The acoustic noise flowing through the acoustic leak between ear and handset has a spectrum which changes as a function of the pressure of the handset against the ear. To evaluate the performance of booths, the acoustic attenuation (L_{RNE}) of this path can be taken into consideration.

The noise N_s is due to sidetone changes according to the telephone set used and it generally has a spectrum which is different from that of N_b. In spite of their mutual correlation, the power summation of the respective spectra seems the best estimate of the global noise (N_g) which affects the ear engaged in the conversation.

In addition, the noises at the two ears (N_a, N_b) are generally different, both in level and in spectrum; experimental intelligibility measurements [1] [2] have demonstrated that this disturbing effect can be evaluated by subtracting 10 dB from the noise level (N_a) at the free ear.

The aforesaid experiment measurements have also shown that the *total* equivalent noise N_T to be used in intelligibility calculations is given by the amplitude sum of noise spectra at the two ears. Consequently, the total equivalent noise N_T is given from the relation:

$$N_T = 20 \log_{10} \left(10^{\frac{N_a - 10}{20}} + 10^{\frac{N_g}{20}} \right) \qquad \text{dB}$$

The sidetone noise N_s is a function of the mouth-to-ear sidetone loss L_{MEST} and it should be measured at the actual noise level, typically 65 dB SPL, under diffuse field conditions. This is particularly important in the case of telephone sets with carbon microphones or of electronic telephone sets with automatic gain control or provided with noise cancelling microphones.

3 Calculation of the booth or acoustic hood efficiency

Given a particular telephone booth or an acoustic hood, the following procedure shall be followed for determining the articulation index in actual operating conditions.

Calculate:

a) the noise spectrum N_a inside the booth by subtracting the acoustic attenuation of the booth (L_a) from the external noise spectrum (N_e). The attenuation should be measured in third octave bands, with a person inside the booth (or a baffle providing an equivalent acoustic absorption) and in a diffuse field condition;

b) the spectrum of the noise N_b by subtracting the leakage attenuation of the handset (L_{RNE}) from the noise spectrum inside the booth N_a;

---

[1] Documentation about the specifications in this Recommendation is not yet sufficient to confirm their validity, thus they are subject to future enhancement and should be regarded as provisional.

c) the sidetone noise spectrum N_s by subtracting the acoustic sidetone attenuation L_{RNST} [2)] from the noise spectrum inside the booth N_a;

d) the spectrum of global noise N_g at the ear pressed against the handset as the power sum of N_s and N_b;

e) the spectrum of total equivalent noise N_T as the amplitude sum of noises at both ears, after having subtracted 10 dB from the noise spectrum at the ear not engaged;

f) the articulation index, AI by Kryter's method [3], assuming a listening speech level of 70 dBA, a value corresponding to the limit of the attenuation of the line loss distribution.

An example of application of the calculation method is shown in Appendix I.

4 Efficiency limits of booths and acoustic hoods

Efficiency of booths or acoustic hoods can be considered satisfactory if an AI equal to 0.6 is guaranteed.

This value corresponds for most languages to a logatom intelligibility of 80% inside the booth, according to the results of French and Steinberg [4], in Figure 1/P.32. It can be assumed as the minimum acceptable limit of performance, corresponding to the maximum external noise level that the booth can withstand in order to guarantee a good quality of telephone transmission inside the booth.

Therefore, each booth can simply be classified by specifying a *maximum external noise level* (MENL), which is the level that gives AI = 0.6.

The MENL that classifies the telephone booth shall be determined by repeating the calculation of the AI, as is indicated in § 3, with different levels of external noise. By means of the curve representing the values of the AI as a function of the outside noise level, the MENL corresponding to an AI = 0.6 can be determined. This MENL depends not only on the acoustic attenuation of the booth or acoustic hood, but also on the received speech level which is assumed to have a reference value of 70 dBA, and on the sidetone performances of the telephone set which should be measured at a proper sound pressure level, (about 65 dB SPL) and in free field conditions.

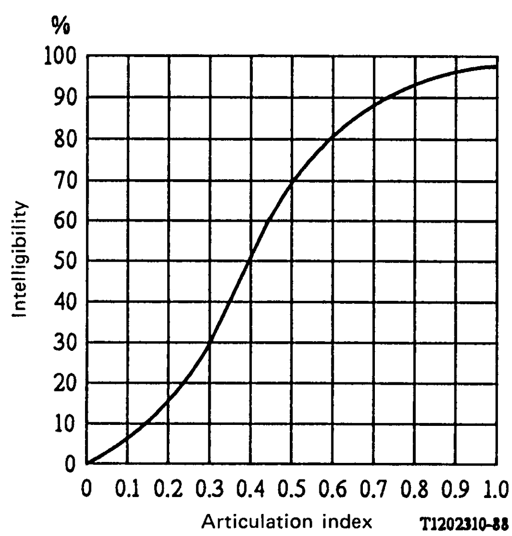


FIGURE 1/P.32

Relation articulation index and logatom intelligibility

---

2) It is important to determine the room noise sidetone sensitivity L_{RNST} which makes use of a diffuse room noise source within the booth. It may also be necessary to include within the booth a manikin to simulate the presence of a subscriber.

5 Speech privacy of telephone communications

The booth can also guarantee speech privacy of conversation by reducing the vocal signals radiated towards outside in order to make them unintelligible. Applying Kryter's calculation method of the articulation index of the speech signals transmitted through the booth to the external ambient at a predetermined noise level, the distance at which the logatom intelligibility or AI falls to a pre-determined value (for example, AI = 0.3) can be estimated. This method can be used to determine the curves of equal intelligibility (isophenes) in any direction, increasing distance from the booth.

Note – The quality improvement of the conversation for the subscriber at the other end of the telephone connection, during a call with a telephone in a booth or acoustic hood has not yet been studied. The evaluation of this aspect is required in any case to consider a number of other factors such as the natural increase of speech loudness in noisy environments and the effective signal-to-noise ratio of transmitted signals.

ANNEX A

(to Recommendation P.32)

Example of efficiency calculation of a telephone booth

The articulation index (AI) is calculated according to Kryter's method.

The acoustic attenuation of a telephone booth measured in an echo chamber at each one-third octave band is reported in Table A-1/P.32, column 2. The total noise level outside the booth is 80 dBA and the sound level of the noise at each centre frequency band is indicated in column 3. The sidetone response characteristics (L_{RNST}) of the telephone set used inside the booth is given in column 4.

The noise level inside the booth at each centre frequency band (N_a) is obtained by subtracting column 2 from column 3 (column 5). It is supposed that the handset of the telephone instrument used in the booth has the acoustic attenuation indicated in Figure A-1/P.32 and reported in column 6.

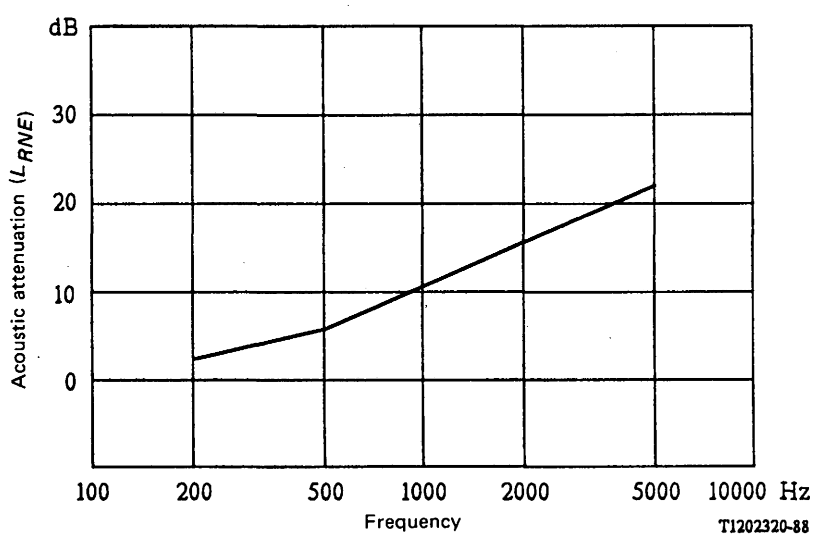


FIGURE A-1/P.32

Acoustic attenuation of handset pressed against the ear

The values of the noise (N_b) due to acoustic leakage between ear and handset obtained by subtracting column 6 from column 5 are reported in column 7.

The values, at each frequency band, of the sidetone noise (N_s) obtained by subtracting column 4 from column 5 are reported in column 8. The global noise at engaged ear (N_g) is reported in column 9 as the power sum of the levels indicated in columns 8 and 7. The total equivalent noise is obtained by adding the levels of column 9 to the values of column 5 reduced by 10 db (column 10). The speech spectrum (β') is reported in column 11 and the signal-to-noise ratio corrected by 12 dB (considering the peaks of the speech signal) is indicated, at each one-third octave band, in column 12. Kryter's coefficients are indicated for each one-third octave band in column 13.

The articulation index (AI) is obtained by multiplying the values of column 12 by those of column 13 and adding the results. By repeating the calculation with other external noise levels, it is possible to draw the diagram of the AI as a function of external noise levels for the considered booth, as shown in Figure A-2/P.32. It can be seen that this booth is designed for withstanding a maximum external noise of about 77 dBA which is the MENL value that classifies the booth.

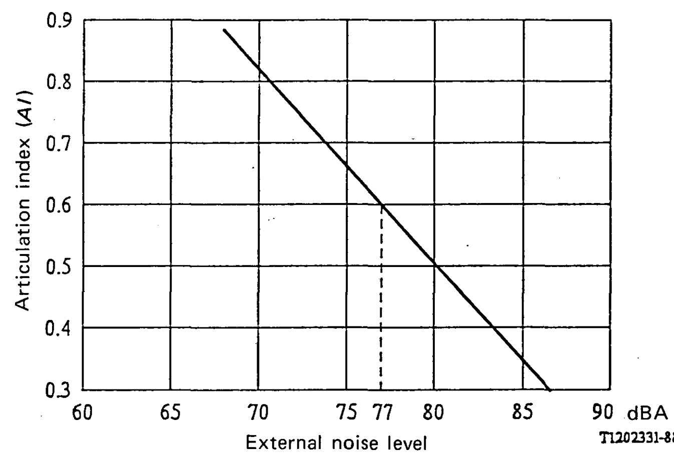


FIGURE A-2/P.32

Articulation index as a function of external noise levels

(MENL of the booth = 77 dBA)

TABLE A-1/P.32

Central frequency one-third octave band (Hz)	Acoustic attenuation of the booth, L_a (dB)	External noise, N_e (dB SPL)	Acoustic sidetone attenuation, L_{RNST} (dB)	Noise inside the booth, N_a (dB SPL)	Acoustic attenuation of handset, L_{RNE} (dB)	Noise due to acoustic leakage, N_b (dB SPL)	Sidetone noise, N_s (dB SPL)	Global noise at engaged ear, N_g (dB SPL)	Total equivalent noise, N_T (dB SPL)	Speech spectrum, β′ (dB SPL)	$\frac{\text{Signal} + 12\text{ dB}}{\text{Noise}}$	Kryter's coefficient	Products (13) × (12)
(1)	(2)	(3)	(4)	(5) =(3)-(2)	(6)	(7) =(5)-(6)	(8) =(5)-(4)	(9)	(10)	(11)	(12) =(11)+12dB-(10)	(13)	(14)
200	10	77.5	12	67.5	3	64.5	55.5	65.0	68.1	61	4.9	0.004	0.0196
250	13	76.5	12	63.5	4	59.5	51.5	60.1	63.4	63	11.6	0.001	0.0116
315	13	73.5	11	60.5	5	55.5	49.5	56.5	60.0	64	16.0	0.001	0.0160
400	15	74.0	9	59.0	6	53.5	50.0	54.8	58.4	65	18.6	0.0014	0.0260
500	14	72.5	9	58.5	7	51.5	49.5	53.6	57.4	65	19.6	0.0014	0.0277
630	14	72.0	10	58.0	8.5	49.5	48.0	51.8	56.1	63	18.9	0.002	0.0378
800	16	72.0	12	56.0	10.0	46.0	44.0	48.1	53.1	62	20.9	0.0020	0.0418
1000	15	71.0	12	56.0	11.5	44.5	44.0	47.3	52.7	61	20.3	0.0024	0.0487
1250	15	69.5	9	54.5	13.0	41.5	45.5	47.0	51.9	60	20.1	0.0030	0.0603
1600	15	68.0	9	53.0	14.5	38.5	44.0	45.1	50.1	58	19.9	0.0037	0.0736
2000	11	66.0	8	55.0	16.0	39.0	47.0	47.6	52.4	54	13.6	0.0037	0.0503
2500	11	64.0	10.5	53.0	17.5	35.5	42.5	43.3	49.2	49	11.8	0.0034	0.0401
3150	12	62.0	14	50.0	19.0	31.0	36.0	37.2	44.7	47	14.3	0.0034	0.0486
4000	12	61.5	14	49.5	20.5	29.0	35.5	36.4	44.1	39	6.9	0.0024	0.0166
TOTAL (dBA)		80.0		66.3					64.7	70.0			AI = 0.52

SPL Sound pressure level

References

[1] CCITT Contribution – COM XII-No. 122, (France), Study period 1981-1984.

[2] CCITT Document – Annex 2, AP VII-No. 115.

[3] KRYTER, (K.): Methods for the calculation and use of Articulation Index, *J.A.S.A.* Vol. 34, 1962.

[4] FRENCH, (N. R.) and STEINBERG (J. C.): Factors governing the intelligibility of speech sounds, *J.A.S.A.* Vol. 19, 1947.

Bibliography

CCITT Contribution – COM XII-No. 139, (Italy), Study period 1973-1976.

CCITT Contribution – COM XII-No. 130, (Norway), Study period 1977-1980.

CCITT Contribution TD 26, (Sweden), WP Laboratory (Geneva, 17-19 January 1984)

KRYTER (K.): The effects of noise on man, *Academic Press*, pp. 70-77, 1970.

Recommendation P.33

SUBSCRIBER TELEPHONE SETS CONTAINING EITHER LOUDSPEAKING RECEIVERS OR MICROPHONES ASSOCIATED WITH AMPLIFIERS

(Mar del Plata, 1968; amended at Geneva, 1972 and 1980)

The CCITT,

considering

(a) that an increasing number of loudspeaker sets is being used in the telephone network,

(b) the complex nature of factors introduced by this equipment and affecting telephone transmission performance,

(c) the need to help Administrations to determine the conditions in which the use of such equipment may be authorized in telephone networks,

makes the following recommendation:

(1) In order to avoid overload of carrier systems, the mean long-term power of speech currents should not exceed the mean absolute power level assumed for system design. In Recommendation G.223 [1] the value adopted for this mean power level is −15 dBm0 (mean power = 31.6 microwatts). Loudspeaker telephones having a sending sensitivity that complies with Recommendation P.34 can be assumed to fulfil this Recommendation. Furthermore, in order to avoid excessive crosstalk from high-level speech currents and/or inadequate received volume from low-level speech currents, care should be taken to ensure that the variation of speech currents is not substantially greater than that from modern handset telephones.

(2) Administrations should take the necessary precautions so that the person listening may be able to break the sending circuit if oscillations occur, or provide for suitable methods so that a device controlled by the voice may prevent oscillations.

Reference

[1] CCITT Recommendation *Assumptions for the calculation of noise on hypothetical reference circuits for telephony*, Vol. III, Rec. G.223.

TRANSMISSION CHARACTERISTICS OF HANDS-FREE TELEPHONES

(Melbourne, 1988)

1 Introduction

The sending and receiving sensitivities of handset telephones, normally expressed as Loudness Rating (LR) values, are used in most countries in connection with their national transmission plan for the design of the national network.

However, since it is possible to fulfil Recommendations such as G.121 by distributing LR values between the telephone sets and the network in different ways, it is not possible to issue an international Recommendation stating LR values of telephone sets alone – whether these are handset or hands-free telephones.

On the other hand, it is possible to recommend sensitivity values for hands-free telephones (HFTs) relative to the standard handset telephone used nationally. The object of such Recommendations should be to obtain equivalent performance with both types of telephones, at least with respect to send and receive loudness. This means that the average user's behaviour and preferences while talking and listening must be taken into account. The relative sensitivities defined in §§ 2 and 3 are derived from performance tests aimed at fulfilling this requirement.

Other important features contributing to the quality of telephone calls made from hands-free telephones cannot presently be dealt with by existing Recommendations and are studied within Question 2/XII [1].

For loudspeaking telephones (see Recommendation P.10) which do not provide full hands-free operation, the relevant parts of this Recommendation may be referred to.

2 Sending sensitivity

The sending LR (SLR) of an HFT should be about 5 dB worse (i.e. higher) than the SLR of the corresponding handset telephone (the actual value will depend on the type of handset used).

Note – Conversation tests in several countries have shown that comparable speech voltages are obtained on the line when the sending loudness rating of the HFT is 5 dB higher than that of the handset telephone used.

The difference of 5 dB has several components:

a) the average talking level for HFTs, which is about 3 dB higher than for handsets;

b) the output level from a handset telephone in conversational use, which is about 1-2 dB lower than what is obtained in the speaking position specified for loudness ratings measurements;

c) other minor differences such as different frequency response curves.

If the sending sensitivity is controlled by the room noise level, this control should be designed to compensate the expected rise of the talking level with room noise.

It should not be possible for the user to adjust the sending sensitivity.

3 Receiving sensitivity

The receiving sensitivity of a hands-free telephone without automatic gain control should be adjustable within a range of 15 to 30 dB. This range should span the value of the receiving loudness rating (RLR) which is equal to that of the corresponding handset telephone, as well as a RLR value about 10 dB better.

Note 1 – Every precaution should be taken to ensure that the increase in gain due to the volume control does not allow the overhearing of other telephone conversations due to crosstalk.

Note 2 – In principle, the RLR of the HFT should be equal to the RLR of the corresponding handset telephone in a quiet room. The range of room noise levels met in normal office use necessitates, however, an additional gain of at least 10 dB.

For hands-free telephones equipped with an automatic gain control for the receive level (the gain being controlled by the incoming speech voltage), loudness ratings may not be applicable. In this case, the HFT should be designed so that the listening level at the maximum line length for which the HFT is intended to be used can be preset to a value that may be considered as the best compromise between the levels required for listening in quiet and noisy rooms.

Note 3 – The preferred listening level depends on the room noise level as well as on other external conditions. There is, furthermore, a great variance between individual listeners.

The average preferred level for listening only appears to be a sound pressure level of about 65 dB for 45 dBA room noise, or 70 dB for 55 dBA room noise. However, to obtain maximum Mean Opinion Scores in conversation tests, listening levels of about 5 to 10 dB higher may be required.

4 Frequency response curves

4.1 *Sending*

Available information indicates that the optimum slope of the sending response curve when measured with the HFT on a table lies between 0 and +3 dB/octave, if the receiving response curve is flat.

Only under highly reverberant conditions may a somewhat higher preemphasis increase the intelligibility. Therefore, if a frequency compensation for the probable cable attenuation is included, the sending curve should not rise with frequency by more than 2-3 dB/octave.

Below 300 Hz there should be a gradual roll-off. The slope may be steeper below 200 Hz.

Note – The interval 200-300 Hz makes a significant contribution to the naturalness of the transmitted speech and should therefore be included in the transmission band of the HFT.

Above 4000 Hz, a roll-off by at least –6 dB/octave (preferably –12 dB/octave) is appropriate in order to avoid interference by crosstalk to adjacent channels in certain types of long-distance circuits.

4.2 *Receiving*

The receiving response curve should be substantially flat in the frequency range of 200-4000 Hz.

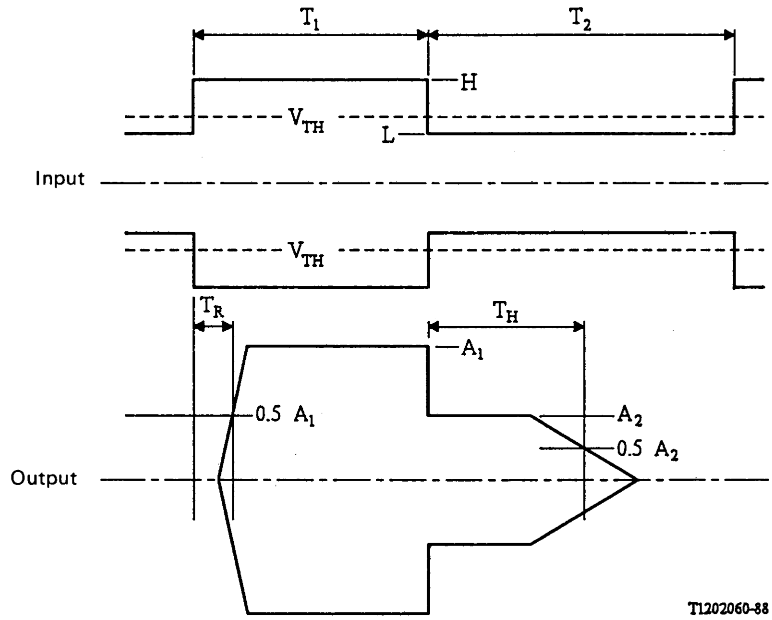
The requirement refers to the sound pressure in the undisturbed field at the listener's position with a set-up including the table as described in § 6.

5 Switching characteristics

Most loudspeaker and hands-free telephones contain voice-switched circuits whose main function is to avoid singing through acoustic feedback. Such circuits insert a loss in either the sending or receiving direction in various ways. Switching from one direction to the other occurs when a signal above a given threshold is applied from the opposite direction, or when the control circuit, taking into account the relative levels and the nature of the signals in both directions, allows the switching.

The fundamental voice-switching parameters of the switching function are defined as follows (see Figures 1/P.34 and 2/P.34):

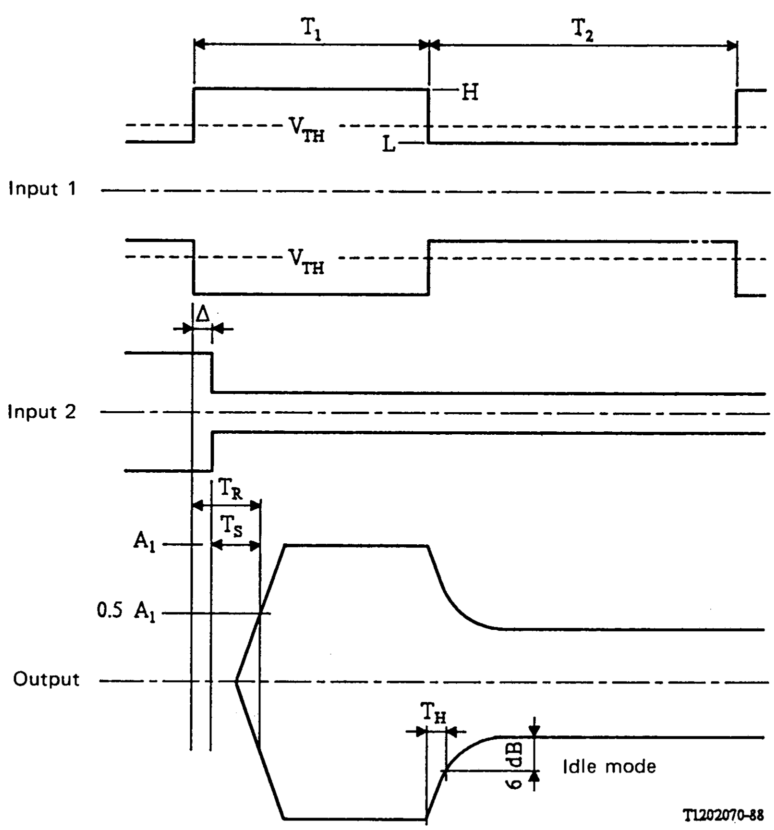
- threshold level V_{TH}: minimum necessary signal level for removing insertion loss,
- build-up time T_R: time from the input signal going above the threshold level up to 50% of the complete removal of the insertion loss,
- hang-over time T_H: time from the input signal going below the threshold level up to the insertion of 50% of the switched loss,
- switching time T_S: time from one transmission direction to the other (see Figure 2/P.34).



T_R Build-up time
T_H Hang-over time
V_{TH} Threshold level

Note – In the case shown here, the switching time T_S is larger than $T_R + T_H$.

FIGURE 1/P.34



T_S Switching time

FIGURE 2/P.34

By a suitable choice of parameter values, the degradation of speech quality that is introduced by voice switching can be made negligible, while an inadequate choice of parameter values, switching times in particular, may lead to serious clipping effects and loss of initial or final consonants in the transmitted speech.

Measurements of voice switching characteristics may be divided into those dealing with:

a) characteristics for alternate conversation, in which two parties communicate by alternating speech spurts without interrupting each other. In this case, it may be assumed that the voice switch circuit returns to an idle state before being activated by an input signal in either direction;

b) characteristics for simultaneous conversation, in which both parties may interrupt each other by simultaneous talk, or where speech at one end of a connection is present simultaneously with noise at the other end.

The first case is of fundamental importance, as its characteristics also affect simultaneous conversation characteristics, and hands-free telephones should therefore always be checked in that respect.

A suitable signal for measuring the characteristics for case a) above consists of a periodic tone burst signal (see Figure 1/P.34). The on/off times T1/T2 and the amplitudes H and L should be adjustable. For case b), in order to switch alternately the hands-free telephone from sending to receiving state, the use of two out-of-phase tone burst sequences is recommended (for instance acoustical 1 kHz, electrical 400 Hz). Switching characteristics measured this way will probably be more readily used in the analysis of subjective conversation tests results.

There are three types of hands-free sets under consideration:

5.1 *Type 1* – Hands-free telephone sets for which switching occurs when an absolute level V_{TH} is reached

In general, it is desirable to keep the threshold value low, the build-up time short and the hang-over time long. On the other hand, in practical applications extremely short build-up times (about a few milliseconds) may cause the voice-switching circuit to be operated by impulsive noises, while very long hang-over times are likely to impede the natural switch-over in conversation. Furthermore, if the threshold level is more than 25 dB below the active speech level, the voice-switching circuit will be activated too easily by ambient noise.

The following switching characteristics are recommended:

a) The build-up time T_R should be less than 15 ms, preferably below 10 ms.

b) The hang-over time T_H should be greater than 100 ms. If the threshold level is in the preferred range, values of T_H between 150 and 250 ms are recommended. Hang-over times greater than 400 ms do not improve the performance noticeably.

c) The threshold level V_{TH} should be at least 20 dB below the active speech level. Levels between -20 and -15 dB may be used if the hang-over time is greater than 300 ms. Levels above -15 dB should not be used.

   In order to measure V_{TH}, the amplitude is gradually increased from a low level until switching occurs. By doing this, an absolute threshold value is obtained. Generally, the threshold is expressed as the difference between this value and the average r.m.s. speech voltage present in the active state.

5.2 *Type 2* – Hands-free telephone sets for which switching depends on the relative levels in both transmission directions, and also in some cases on noise levels (acoustical and electrical), amplifiers gains, automatic gain controls, previous transmission direction, etc.

The following values are recommended:

a) T_R should be less than 15 ms, preferably below 10 ms,

b) T_H can be less than 50 ms,

c) T_S is recommended to be approximately 100 ms and is measured by using 2 excitation signals (see Input 1 and Input 2 in Figure 2/P.34).

   Note – Under highly reverberant conditions, some hands-free sets with such a T_S may operate in an unsatisfactory way.

More information about measuring levels and methods can be found in the Handbook on Telephonometry in § 3.5.

5.3 *Type 3* – Hands-free sets using echo cancellation techniques

Some indications about the evaluation of sets using echo cancellation are given in Recommendation P.30.

Note – For loudspeaking telephone sets, an insertion loss may be introduced in the receiving side to avoid the acoustical coupling with the handset microphone. This insertion loss may be introduced when the received level on the loudspeaker is too high, or when the signal from the handset microphone is sent onto the loudspeaker at too high a level.

It is recommended that the delay of application and withdrawal of this insertion loss be limited to 20 ms and its value limited to avoid any clipping effect on the received speech.

6 Conditions of measurement

For both subjective and objective measurements, physical test arrangements as described in this section should be used.

6.1 *Test table*

During the measurements, the HFT is placed on a table defined as follows:

The surface of the table should be hard (e.g. polished marine plywood or suitable hardwood), flat, rigid and horizontal to provide a sound-reflecting surface on which the HFT being tested rests. The dimensions of the table should be such that the surface area is about 1 m^2 but not less than 0.96 m^2 and the width not less than 800 mm [2].

Note – This arrangement should be used for all measurements, including the recording of frequency responses, although diffraction effects due to the table are likely to cause severe dips or peaks in the response curve (see § 6.5.2).

6.2 *Test arrangements*

The physical test arrangements of one- and two-piece HFTs [3] for subjective and objective measurements is shown in Figure 3/P.34.

If the projections of the housing are not rectangular, the point B is positioned at the crossing of the centre line through the housing and the outline of the vertical projection of the housing.

The edge of the front of the box should be perpendicular to the line A-B.

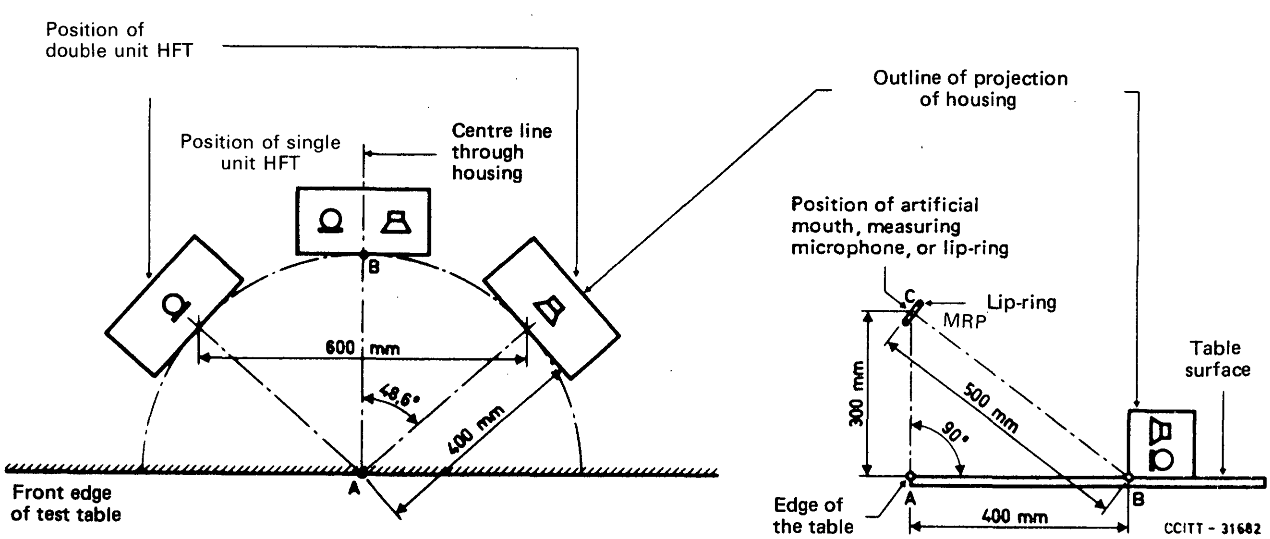


FIGURE 3/P.34

Physical test arrangements for subjective and objective measurements

6.3 *Test environment*

When performing tests, the room acoustics must not have a dominating influence. It is recommended for objective measurements that the test environment be practically free-field (anechoic) down to a lowest frequency of 175 Hz, and be such that the test arrangement lies totally within the free-field volume.

Note – Satisfactory free-field conditions may be considered to exist where errors due to the departure from ideal conditions do not exceed $\pm$ 1 dB.

The tests should be made in an environment where the ambient noise level is negligible. For objective measurements this is achieved if the Noise Rating (NR) of the Noise Criterion (NC) is lower than 15 [4, 5]. For subjective tests, it may be sufficient to keep the sound level of ambient noise below 35 dBA.

6.4 *Subjective determinations*

Loudness rating should be determined in accordance with Recommendation P.78.

Note – Some information about reference equivalents can be found in the CCITT *Red Book* (Vol. V 1985), or in the Handbook on Telephonometry.

6.4.1 *Sending*

The talking level for the measurement of sending loudness rating (SLR) of an HFT should normally be the same as specified for measurements on handset telephones.

It is not necessary for the talker during the test to shift between the reference microphone guard-ring and the guard-ring positioned relative to the HFT if the obstacle effect of the reference microphone can be assumed to be negligible.

Normally the specified talking level and the use of a conventional test phrase or sentence should be sufficient to ensure that a voice-switched HFT will be in the sending condition during the determination of SLR. If this is not the case the talking level may be increased by up to 5 dB, which may be compensated in the reference system to preserve the same listening level.

If the sending sensitivity is controlled by the room noise level the subjective measurement should be done in a quiet environment (< 35 dBA). Further information about the HFT performance may then be estimated by repeating the sending measurements with increasing levels of room noise, up to a maximum of 60 dBA.

6.4.2 *Receiving*

The talking level at the reference microphone for the measurement of RR25E or RLR should normally be the same as specified for the measurement of handset telephones. This should normally ensure that when loudness balance is achieved between the reference system and the test system path, a signal of sufficient magnitude is present at the HFT to switch it into the receive condition.

Problems can sometimes occur when approaching the balance condition from the condition of high attenuation in the balance attenuators, when the low level input signal may fail to switch the HFT into the receiving condition. If this does occur the talking level may be increased by up to 5 dB in order to minimize the difference in loudness.

Note – The listening level will thus also increase at balance, but in this case it will not be possible to correct it by changing the reference system attenuator.

Obtaining the loudness balance for the receiving condition may be facilitated by use of a loudspeaking intermediate reference system. The specification of such a system is, however, outside the scope of this Recommendation.

6.5 *Objective evaluations*

Objective evaluations of loudspeaker and hands-free telephones concern:

- the sending and receiving frequency sensitivity curves measurements,
- the objective determination of loudness ratings according to the method described in Recommendation P.79.

Note – Other methods for calculating loudness ratings used by some Administrations for their own internal planning purposes can be found in Supplement No. 19.

6.5.1 *Sensitivity measurements*

6.5.1.1 *Sending sensitivity measurements*

The sending response curves of a looudspeaker and/or hands-free telephone is recorded at the output terminals of the thetelephone with the same electrical connections as for the handset telephones. The acoustical input to the telephone microphone is supplied from an artificial mouth in the position shown in Figure 3/P.34.

In such a case, the sending sensitivity of the local telephone system is expressed as follows:

$$S_{mJ} = 20 \log_{10} \frac{V_s}{p_m} \qquad \text{dB rel 1 V/Pa}$$

where V_s is the voltage across a 600 ohm termination and P_m is the sound pressure at the MRP.

The measuring level proposed in Recommendation P.64 may be used: −4.7 dBPa at the MRP (Figure 3/P.34), which corresponds to −29 dBPa at 50 cm from the lip when there is no table nor set.

Note 1 – Some HFTs use "noise guard" circuits and therefore the source signal has to be modified. A suitable method is to pulse the source signal at an adequate rate, e.g. 250 ms "ON" and 150 ms "OFF". Sending sensitivities determined for HFTs in this way are not suitable for use in calculating send loudness ratings (SLR). For this purpose, the reference sound pressure should have a level at the MRP which is (on average, over the frequency range of interest) 24.2 dB higher.

6.5.1.2 *Receiving frequency sensitivity response measurements*

The receiving sensitivity of a loudspeaker and/or hands-free telephone is expressed as follows:

$$S_{Je} = 20 \log_{10} \frac{p_R}{\frac{1}{2} E_J} \qquad \text{dB rel Pa/V}$$

where p_R is the sound pressure at point C in Figure 3/P.34 and E_J is the e.m.f in the 600 ohms source.

6.5.2 *Measure and computation of loudness ratings*

6.5.2.1 *Sending*

The computation of the sending loudness rating may be performed according to Recommendation P.79 by using the frequency sensitivity response measured between the electrical output of the set and the acoustical sound pressure at the MRP (Figure 3/P.34).

Note – Other methods for calculating loudness ratings used by some Administrations for their own internal planning purposes can be found in Supplement No. 19.

However, some care must be taken in the test design and the interpretation of the results. Results available up to now concern only a limited number of sets and the measuring signal is of some importance. Under some conditions, an artificial speech-like signal may activate the noise-guard circuits (by inserting some loss at the sending side). Better results are expected by using an artificial voice satisfying Recommendation P.50 (temporal characteristics of the signal closer to those of real speech).

6.5.2.2 *Receiving*

Objective measurements described in § 6.5.1.2 are made with a free-field microphone at point C (Figure 3/P.34).

Loudness Ratings are computed following Recommendation P.79, provided the following phenomena are taken into account:

- the diffraction effect of the listener head,
- an appropriate correction for the difference between one-ear and two-ears listening.

These subjects are still under study under Question 2/XII.

Provisionally, a correction term of 14 dB should be subtracted in the computed loudness ratings.

Note – Other methods for calculating loudness ratings used by some Administrations for their own internal planning purposes can be found in Supplement No. 19.

References

[1] CCITT – Question 2/XII Contribution COM XII-No. 1, Study Period 1989-1992.

[2] European Committee for Standardization (CEN) *Office chair/desk working position – dimensions and design requirements*, CEN: prEN91/August 1981.

[3] CCITT *A method for measuring the sensitivity of a loudspeaking telephone set*, Annex 2 to Question 17/XII, White Book, Vol. V, ITU, Geneva, 1969.

[4] ISO *Assessment of noise with respect to community response*, ISO Recommendation 1996, 1971.

[5] BERANEK (L. L): Noise and Vibration Control, *McGraw Hill*, pp. 564-566, New York, 1971.

Recommendation P.35

HANDSET TELEPHONES

(Malaga-Torremolinos, 1984; amended at Melbourne, 1988)

1 Transmission characteristics

The transmission plan for international telephone connections is given in Recommendation G.101.

Recommendations G.111 and G.121 deal with the transmission quality, i.e. loudness ratings for international and national telephone connections, respectively.

These Recommendations permit administrations to split the requirements between analogue telephone sets and the other parts of the network as long as the overall specifications are fulfilled.

Therefore no precise specifications can be given for analogue telephone sets, although some design considerations can be provided. The latter are contained in Supplement No. 10.

Recommendations for digital telephones are found in Recommendation P.31.

2 Handset dimensions

The shape and the dimension of the handset have an important influence on both send and receive levels. The earpiece must be capable of forming a good seal to the ear and the handgrip of the handset must be such that it will encourage the user to hold it to the head in the optimum position.

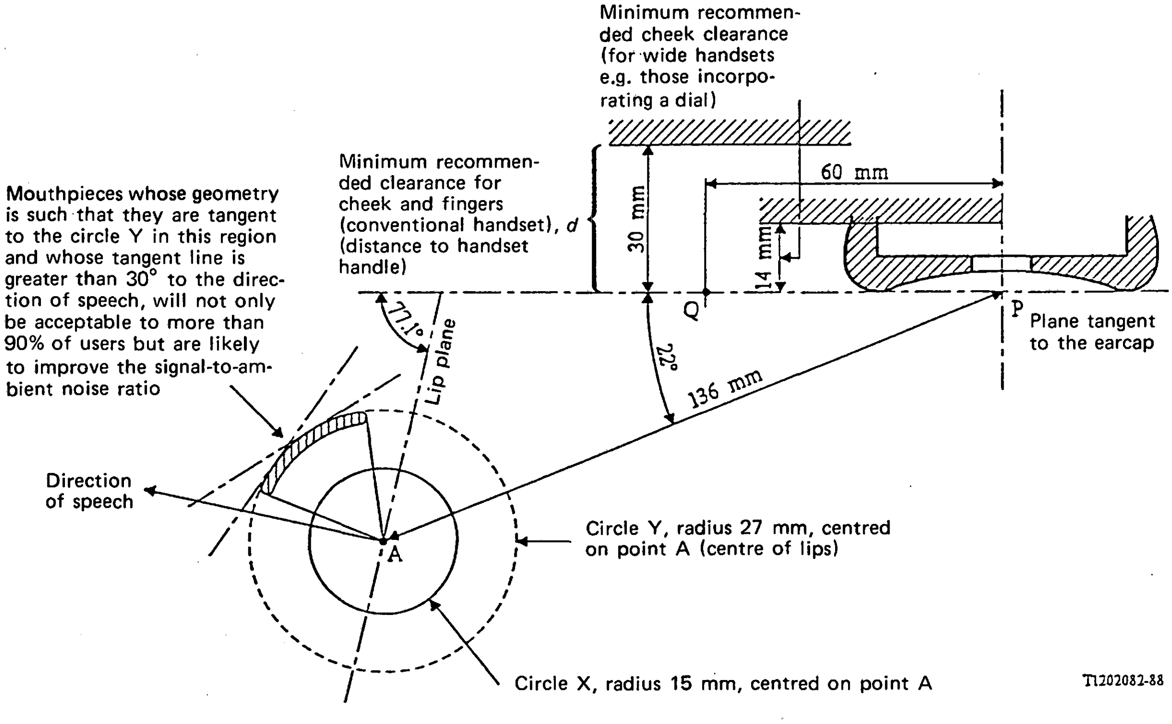
Reference [1] is an ergonomic study which presents data on the distribution of the relevant finger and head dimensions.

A later head dimension study carried out in the People's Republic of China is reported in [2]. A subsequent investigation [3] shows that, for convenience in use, the mouthpiece of the handset should be somewhat outside (e.g. 10-12 mm) a circle enclosing the centre of the lip of 80% of the subjects tested (over 4000). A handset conforming to these dimensions (see Figure 1/P.35) will then be acceptable to more than 90% of users. When a longer lip-to-mouthpiece distance is chosen, the signal-to-ambient-noise ratio will be worse and recommended LSTR values will be more difficult to meet (see Recommendations G.121, P.11, P.76, P.79 and Supplement No. 11). Therefore both signal-to-ambient-noise ratio and mouthpiece position for convenient use must be considered and probably a compromise must be made.

3 Recommendation on handset

Based on the information given above, the CCITT recommends that handset telephones conform to the dimensions outlined in Figure 1/P.35, with respect to mouthpiece positions and cheek-to-handset clearance.

Note – An earpiece with a design that forms a good seal to the IEC 318 ear (Recommendation P.51) will facilitate testing both in laboratories and during manufacturing. Experience has shown that earpieces with a good seal to the IEC 318 artificial ear also give in most cases a good seal to the human ear.



Note 1 – Point A is located at the centre of circle X which enclosed 80% of lip positions of the subjects tested [2], [3].

Note 2 – The circle Y encloses the ellipse given in [1].

FIGURE 1/P.35

References

[1] CCITT – Contribution COM XII-No. 49 (ITT), Study Period 1973-1976.

[2] CCITT – Contribution COM XII-No. 21 (People's Republic of China), Study Period 1977-1980.

[3] CCITT – Contribution COM XII-No. 112 (People's Republic of China), Study Period 1977-1980.

Bibliography

CCITT – Contribution COM XII-No. 32 (U.K), Study Period 1973-1976.

EFFICIENCY OF DEVICES FOR PREVENTING THE OCCURRENCE OF EXCESSIVE ACOUSTIC PRESSURE BY TELEPHONE RECEIVERS

(Malaga-Torremolinos, 1984; amended at Melbourne, 1988)

The use of devices for preventing the occurrence of excessive acoustic pressure by telephone receivers is recommended in Recommendation K.7. Methods for checking the efficiency of such devices in response to short duration impulses and for longer duration disturbances, such as tones, are given in this Recommendation. A method is also given for checking that such devices do not have adverse effects on normal speech signals.

Preliminary Note – On the basis of the findings of scientific studies, several authors or organizations have proposed ear-damage risk criteria based on variations in acoustic pressure, under impulse conditions for which, parenthetically, there is no single definition. Likewise, ear-damage risk criteria have also been proposed for longer duration acoustic disturbances, such as tones. However, these criteria cannot be directly transposed to the test conditions and measurements described below. Nor could the results be cross-checked without introducing certain hypotheses that are not specified in this Recommendation, the purpose of which is merely to describe a method simple both in its application and in the analysis of the results obtained. The criteria recommended are based on experience gained in several countries about the telephone receiver quality necessary to ensure the safety of users and operators. Administrations may wish to adopt lower limiting levels to reduce user annoyance caused by acoustic disturbances, but the limiting levels should not be so low as to have adverse effects on normal speech levels.

1 Efficiency of protection against short duration impulses

In order to check whether a telephone set affords satisfactory protection against the risk of acoustic shocks due to short duration impulses, it is recommended that its characteristics be examined as follows:

a) the entire telephone set, including the protective device, is placed in normal operating conditions as regards current supply and its position for the exchange of a call (e.g. with the handset raised);

b) the earpiece of the handset earphone is applied in the normal way to an artificial ear conforming to Recommendation P.51 (which corresponds to IEC Publication 318);

c) the artificial ear is electrically connected to a precision sound level meter conforming to IEC Publication 651, correctly calibrated and having the necessary circuits for measuring peak acoustic pressure levels. This equipment must be of class 2 for prototype testing, and may be of class 3 for checking mass-produced sets;

d) electrical impulses are applied to the telephone set by a suitable assembly which enables these impulses to be superimposed on the d.c. supply without the latter short-circuiting them. These impulses are produced by a generator which conforms with Figure 1/K.17, and whose components are those described for symmetric-pair repeater tests (R_3 = 25 ohms, C_2 = 0.2 μF, see Table 1/K.17). The test voltage is between 0 and 1.5 kV;

e) the telephone set is also checked for self-generated acoustic impulses such as those produced by operation of the hook switch or by dialing;

f) for both cases d) and e) above, the peak acoustic pressure level observed (maximum instantaneous value) should be below 140 dB relative to 20 μPa. In the long term, Administrations are recommended to limit this value to 135 dB for sets in common use.

Note – Administrations may deem it appropriate to use different limits for specific cases, for instance for the headsets used by operators.

2 Efficiency of protection against longer duration disturbances

In order to check whether a telephone set affords satisfactory protection against the risk of acoustic hazards due to longer duration disturbances, such as tones, it is recommended that its characteristics be examined as follows:

a) the entire telephone set, including the protective device, is placed in normal operating conditions as regards current supply and its position for the exchange of a call (e.g. with the handset raised);

b) the earpiece of the handset earphone is applied in the normal way to an artificial ear conforming to Recommendation P.51 (wich corresponds to IEC Publication 318);

c) the artificial ear is electrically connected to a precision sound level meter conforming to IEC Publication 651, correctly calibrated to measure A-weighted sound pressure levels. This equipment must be of class 2 for prototype testing, and may be of class 3 for checking mass-produced sets;

d) a 1000 ± 20 Hz[1)] sinewave signal is applied to the telephone set and its amplitude is increased until it reaches 10 V_{rms} across the set's terminals or until the steady-state acoustic output from the telephone receiver reaches its limiting value, whichever occurs first;

e) the telephone set is also checked for self-generated acoustic disturbances, such as tone dialing signals fed back to the receiver;

f) for both cases d) and e) above, the steady-state A-weighted sound pressure level should be below 125 dBA ("slow" response).

Note 1 – Tones or other disturbances which are inherently limited to less than 0.5 s duration should be evaluated as short duration impulses under § 1. Repetitive disturbances, such as those which might be produced during automatic tone-type dialing, should be evaluated under § 2 using the sound level meter set for "slow" response averaging.

Note 2 – Administrations may deem it appropriate to use different levels for specific cases, for instance, for the headsets used by operators.

3 Effect on normal speech signals

It is recommended to check whether the strong-signal attenuation obtained by protective devices does not cause deterioration of the normal speech signals, e.g. by nonlinear distortion. This may be done by conducting a series of measurements using steady-state sine wave signals at a frequency of 1000 ± 20 Hz and relating to the following magnitudes:

N is an electric voltage level at the terminals of the set. N is determined by the relation:

$$N = 20 \log_{10} \frac{V_{rms}}{0.775} \qquad \text{(dB)}$$

where V_{rms} represents the r.m.s. value of the voltage across the terminals. The value of $V_{rms} = 0.775$ volts (-2.2 dBV) gives $N = 0$ and corresponds to a power level of 0 dBm into 600 ohms.

$P(N)$ is an acoustic pressure produced by the telephone receiver under given conditions, (this may be the pressure measured on an artificial ear in accordance with Recommendation P.51), corresponding to the application of voltage level N across the terminals of the set.

$A(N)$ is an attenuation of electroacoustic efficiency in relation to its reference value determined for $N = -20$ dB. $A(N)$ is determined by the relation:

$$A(N) = 20 \log_{10} \frac{P(-20)}{P(N)} + N + 20 \qquad \text{(dB)}$$

[$A(N) = 0$ when $N = -20$ dB].

The values obtained for $A(N)$ must match those in Table 1/P.36 which have been obtained from measurements carried out on several types of set fitted with various protective devices.

---

1) The ISO list of preferred frequencies includes 1000 Hz. It is a commonly used reference frequency in acoustic testing. Recommendation O.6 suggests 1020 Hz be used when testing PCM systems to avoid being at a submultiple of the 8000 Hz sampling rate. This Recommendation may need to be considered when testing digital telephones.

Note 1 – It may be useful to make a few additional measurements to ensure that, at frequencies between 200 Hz and 4000 Hz, the values observed for $A(N)$ are of the same order.

Note 2 – Some sets of recent design have special features, such as electroacoustic sensitivity which depends on the conditions of d.c. current supply or on the level of the speech signals received, quite apart from the effect of the protective devices. In that case, Administrations intending to use such sets may have to adapt the above conditions, taking care nevertheless to comply with their principles.

TABLE 1/P.36

N (dB)	$A(N)$ (dB)
-20	0
-10	< 0.5
0	$\leqslant 2$

Recommendation P.37

MAGNETIC FIELD STRENGTH AROUND THE EARCAP OF TELEPHONE HANDSETS WHICH PROVIDE FOR COUPLING TO HEARING AIDS

(Malaga-Torremolinos, 1984; amended at Melbourne, 1988)

1 Introduction

Magnetic induction systems incorporated in telephone handsets generate an alternating magnetic field with spatial characteristics which make the field detectable by hearing aids equipped with induction pick-up coils.

Reception of an audio-frequency signal via an induction pick-up coil can often allow an acceptable signal-to-noise ratio to be achieved in cases where the acoustical reception would otherwise be degraded by reverberation and background noise.

The magnetic field strength which enables induction pick-up coils in hearing aids to function effectively must be high enough to produce an acceptable signal-to-noise ratio but not so high as to cause overloading of the hearing aid.

The value of magnetic field strength recommended in this standard has been chosen so that these requirements are met as far as possible.

2 Scope

This Recommendation applies to telephone handsets which provide a magnetic field for coupling to hearing aids. It specifies the level linearity and frequency dependence of the magnetic field strength produced by the handset and characteristics for a calibrated probe coil.

3 Explanation of terms

3.1 *Level of magnetic field strength*

The maximum value of the magnetic field strength is measured in accordance with § 4.2.

3.2 *Plane of measurement*

A plane parallel to the earcap plane at a distance of 10 mm.

4 Magnetic field strength measurements and recommended values

4.1 *Calibration of acoustic receive level*

Using the measurement configuration shown in Figure 3/P.64, the drive level of the oscillator shall be adjusted to produce a sound pressure level of 80 dB at 1000 Hz. This drive level shall be used for measuring the level and frequency characteristics of the magnetic field strength.

4.2 *Magnetic field strength level*

Place (per § 5) the centre of the calibrated probe coil in the plane of measurement and circuit orientate it for maximum coupling. Determine the magnetic field strength at 1000 Hz using the drive level as per § 4.1.

Recommended range of values for the magnetic field strength is:

-17 to -30 dB relative to 1 A/m.

Note – Hearing aids with magnetic pick-up coils primarily intended for coupling to magnetic loops in auditoria in accordance with IEC Publication 118-4 are likely to have a sensitivity that corresponds to a field strength in the upper end of the range recommended for coupling to telephones.

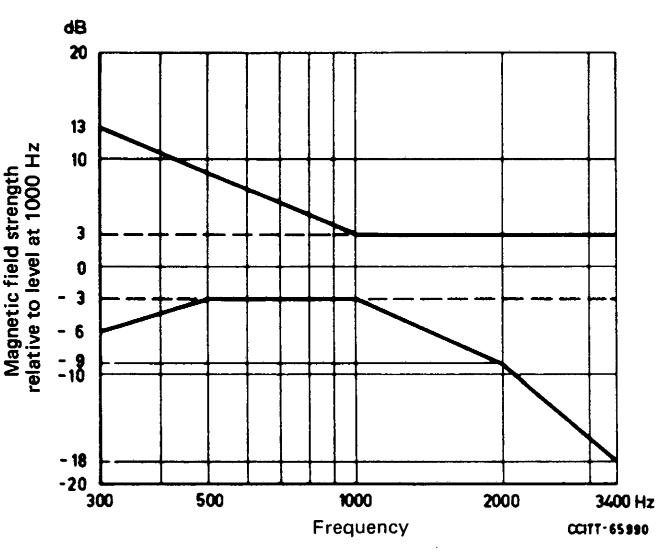
4.3 *Linearity of the magnetic field strength*

With the probe coil positioned as in § 4.2, increase the 1000 Hz drive level specified in § 4.1 by 20 dB and measure the resulting magnetic field strength.

The field strength should increase by 20 dB ± 1 dB, or if the telephone set has a higher linearity the linearity of the magnetic field shall be equally as good.

4.4 *Measurement of frequency characteristics*

With the probe coil positioned as in § 4.2 and the drive level as in § 4.1, vary the frequency from 300 Hz to 3400 Hz and measure the resulting field strength. The magnetic field strength frequency characteristics shall fit within the template shown in Figure 1/P.37.



Note – Preferred frequency characteristics are within the dotted lines (± 3 dB). Range of acceptable characteristics is within the solid lines.

FIGURE 1/P.37

Magnetic field strength frequency characteristics

5 Probe coil

5.1 *Dimensions*

For measuring the magnetic field strength, a calibrated probe coil having the following dimensions is recommended: .

Core: length (12.5 ± 1 mm)
cross section (1 mm ± 0.5 mm) × (2 mm ± 0.5 mm)

Winding: length (10 mm ± 1 mm)
cross section (2 mm ± 0.5 mm) × (3 mm ± 0.5 mm)

The winding shall be shorter than the core.

Note 1 – The magnetic field may be inhomogeneous within distances comparable to the length of the probe coil. The introduction of a magnetic core material may also redirect the magnetic field contours, therefore the magnetic material of the core may be of importance.

Note 2 – The probe coil may be combined with frequency correcting elements to obtain a flat frequency response in the range of 300 Hz to 3400 Hz.

5.2 *Calibration of the probe coil*

The sensitivity as a function of frequency of the probe coil shall be measured with an accuracy of ± 0.5 dB.

A method of producing a homogeneous magnetic field of known intensity is given in IEC Publication 118-1. The harmonic distortion of the magnetic field used for the calibration shall be less than 1%.

5.3 *Distortion*

The distortion of the probe coil shall be less than 2%, when measuring field strength up to +2 dB relative to 1 A/m.

ANNEX A

(to Recommendation P.37)

Measurement of an acousto-magnetic adapter generating a magnetic field

A.1 *Scope*

This annex specifies the measuring method for an acousto-magnetic adapter that converts the acoustic output of an associated telephone receiver to a magnetic field, in accordance with §§ 4.1 and 4.2, that can be received by the magnetic pick-up coil in a hearing aid.

A.2 *Definition of the adapter plane*

The adapter plane is defined as the plane formed by the contacting points of a flat surface against the surface of the acousto-magnetic adapter opposite the earcap connection.

A.3 *Definition of the plane of measurement*

The plane of measurement is defined as a plane parallel to the adapter plane at a distance of 10 mm.

A.4 *Measurement procedures*

Measurements are made in accordance with this Recommendation.

The output sound pressure level of the telephone receiver is measured against the artificial ear without the acousto-magnetic adapter being mounted.

The characteristics of the magnetic field of the acousto-magnetic adapter are measured when mounted on the actual telephone receiver.

Note – In reporting results, the type of telephone set used should be specified.

A.5 *Magnetic field requirements*

The magnetic field produced by the adapter when fitted to a handset should meet the level and frequency characteristic requirements given in §§ 4.2 and 4.4.

A.6 *Physical properties*

Desirable physical properties of the acousto-magnetic adapter are:

- easy to place on the earcap and remove again;
- a firm contact to the earcap so that the acousto-magnetic adapter and the telephone handset can be used as an integrated unit;
- forming a good and well-defined acoustic coupling to the earcap (see Note);
- the surface of the acousto-magnetic adapter defining the adapter plane should be flat or should have a shape easily defining the adapter plane;
- the adapter plane should be approximately parallel to the earcap plane;
- the magnetic field produced by the adapter should be orientated so that the magnetic coupling to the hearing aid is only to a small extent dependent on the position of the hearing aid.

Note – The inner diameter of an acoustic seal is recommended to be equal to the edge diameter of the IEC 318 artificial ear.

Bibliography

Methods of measurement of electro-acoustical characteristics of hearing aids. Part 4: Magnetic field strength in audio-frequency induction loops for hearing aid purposes, IEC Publication 118-4, 1981.

AHLBORG (H.): Speech levels in the Swedish telephone network. *TELE Engl. Ed.*, No. 1, 1978.

DAHLGAARD (T.) and NIELSEN (A. K.): A statistical analysis of speech signals in a local exchange, and a calculation of the line impedance from the natural speech signals. *Teleteknik*, No. 2, 1974.

GLEISS (N.): Preferred listening levels in telephony. *TELE Engl. Ed.*, No. 2, 1974.

Recommendation P.38

TRANSMISSION CHARACTERISTICS OF OPERATOR TELEPHONE SYSTEMS (OTS)

The measurement methods adopted for measuring on operator telephone systems (OTS) which comprise a headset, feeding circuit and subscriber's line (the same principles can be applied to any system that uses a headset), conform to the methods described in Recommendation P.64 with the following exceptions:

1 Sending sensitivities of OTS

In principle, the OTS is similar to the Local Telephone System (LTS) of Recommendation P.64 with the exception that in a headset the earphone and microphone may not have a fixed relationship as has a conventional telephone handset. Those headsets which are not adjustable in distance from the receiver to the microphone should be positioned per Annex A of Recommendations P.76. For those which are adjustable, a modal position of the input port of the mouthpiece must be specified by the manufacturer in 3-dimensional coordinates relative to the lip, horizontal and vertical reference planes of the mouth as defined in Recommendation P.51.

This modal position is defined by the manufacturer to be representative of the position of normal usage.

Note – The term "corner of the mouth" used by some manufacturers in defining the normal use position is assumed to be 21 mm from the centre 9 mm behind the lip plane. The sound field of the artificial mouth is not defined behind the lip plane and therefore measurement points behind the lip plane are not recommended.

The sending sensitivity is then determined as per §§ 2, 4 and 6 of Recommendation P.64

The sending loudness rating (SLR) is computed as described in Recommendation P.79.

2 Receiving sensitivities of OTS

2.1 For headsets using supra-aural earphones, the IEC 318 artificial ear is used.

The receiving sensitivity is determined as per §§ 3, 5 and 7 of Recommendation P.64.

The receiving loudness rating (RLR) is computed as described in Recommendation P.79 using the L_E values of Table 4/P.79.

2.2 For headsets using insert type receivers, the IEC 711 ear simulator is used.

The receiving sensitivity is determined as per §§ 3, 5 and 7 of Recommendation P.64.

2.3 The receiving sensitivity suitable for use in the calculation of loudness requires:

a) a transfer function (S_{DE}) for the eardrum to the ear reference point (ERP) and is given in Table 1/P.51, and

b) the $L_{E(I)}$ values of Table 1/P.38 appropriate for insert type receivers.

The sensitivity is defined as:

$$S_{JE} = S_{Jd} + S_{DE} - L_{E(I)}$$

where

S_{JE} is the sensitivity from the junction to the real ear.

S_{Jd} is the sensitivity from the junction to the 711 IEC ear simulator (eardrum).

$L_{E(I)}$ is the ear coupling loss of insert type receivers (Table 1/P.38).

S_{DE} is the transfer function from the eardrum to the ERP (Table 1/P.51).

The receiving loudness rating (RLR) is computed as described in Recommendation P.79 using S_{JE} from the above formula. (*Note* – The L_E values of Table 4/P.79 have been replaced by the values Table 1/P.38.)

Note 1 – Study is still continuing under Question 8/XII, to evaluate intra-concha, circum-aural and non-contact types of earpieces.

Note 2 – Further information on the measurement of OTS can be found in § 3.4 of the Handbook on Telephonometry.

TABLE 1/P.38

Values of $L_{E\ (I)}$ for insert type receivers

Frequency (Hz)	$L_{E\ (I)}$ (dB)
200	23.0
250	19.0
315	18.0
400	17.4
500	12.8
630	9.0
800	6.8
1000	3.2
1250	1.5
1600	1.4
2000	0.4
2500	-1.5
3150	3.0
4000	0.0

PAGE INTENTIONALLY LEFT BLANK

PAGE LAISSEE EN BLANC INTENTIONNELLEMENT

SECTION 3

TRANSMISSION STANDARDS

Recommendation P.48

SPECIFICATION FOR AN INTERMEDIATE REFERENCE SYSTEM

(Geneva, 1976; amended at Geneva, 1980, Malaga-Torremolinos, 1984, Melbourne, 1988)

Summary

This Recommendation intends to specify the intermediate reference system (IRS) to be used for defining loudness ratings. The description should be sufficient to enable equipment having the required characteristics to be reproduced in different laboratories and maintained to standardized performance.

1 Design objectives

The chief requirements to be satisfied for an intermediate reference system to be used for tests carried out on handset telephones[1] are as follows:

a) the circuit must be stable and specifiable in its electrical and electro-acoustic performance. The calibration of the equipment should be traceable to national standards;

b) the circuit components that are seen and touched by the subjects should be similar in appearance and "feel" to normal types of subscribers' equipment;

c) the sending and receiving parts should have frequency bandwidths and response shapes standardized to represent commercial telephone circuits;

d) the system should include a junction which should provide facilities for the insertion of loss, and other circuit elements such as filters or equalizers;

e) the system should be capable of being set up and maintained with relatively simple test equipment.

Note – The requirements of a) to d) have been met in the initial design of the IRS by basing the sending and receiving frequency responses on the mean characteristics of a large number of commercial telephone circuits and confining the bandwidths to the nominal range 300-3400 Hz.

[1] For other types of telephone, e.g. headset or loudspeaking telephone, a different IRS will be required. The IRS is specified for the range 100-5000 Hz. The nominal range 300-3400 Hz specified is intended to be consistent with the nominal 4 kHz spacing of FDM systems, and should not be interpreted as restricting improvements in transmission quality which might be obtained by extending the transmitted frequency bandwidth.

Since the detailed design of an IRS may vary between different Administrations, the following specification defines only those essential characteristics required to ensure standardization of the performance of the IRS.

The principles of the IRS are described and its nominal sensitivities are given in §§ 2, 3, 4 and 5 below; requirements concerning stability, tolerances, noise limits, crosstalk and distortion are dealt with in §§ 6 to 9 below. Some information concerning secondary characteristics is given in § 10 below.

Certain information concerning installation and maintenance are given in [1].

2 Use of the IRS

The basic elements of the IRS comprise:

a) the sending part,

b) the receiving part,

c) the junction.

When one example each of a), b) and c) are assembled, calibrated and interconnected, a reference (unidirectional) speech path is formed, as shown in Figure 1/P.48. For performing loudness rating determinations, suitable switching facilities are also required to allow the reference sending and receiving parts to be interchanged with their commercial counterparts.

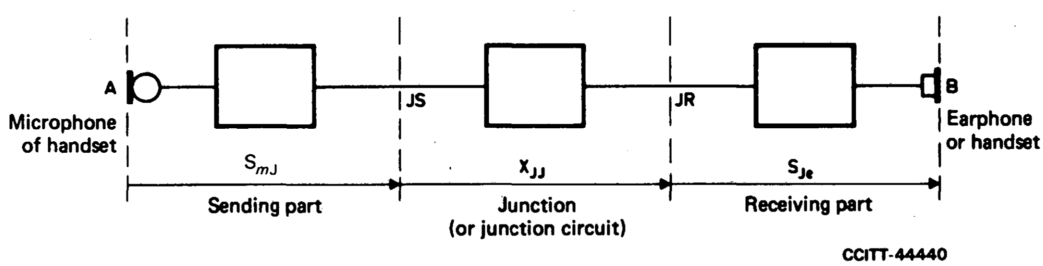


FIGURE 1/P.48

Composition of the complete intermediate reference system

3 Physical characteristics of handsets

The sending and receiving parts of an IRS shall each include a handset symmetrical about its longitudinal place and the profile produced by a section through this plane should, for the sake of standardization, conform to the dimensions indicated in Figure 1/P.35. In practice, any convenient form may be considered use being made, for example, of handsets of the same type as those used by an Administration in its own network. The general shape of the complete handset shall be such that, in normal use, the position of the earcap on the ear shall be as definite as possible, and not subject to excessive variation.

The microphone capsule, when placed in the handset, shall be capable of calibration in accordance with the method described in Recommendation P.64. The earcap shall be such that it can be sealed on the circular knife-edge of the IEC/CCITT artificial ear for calibration in accordance with IEC 318, and the contour of the earcap shall be suitable for defining the ear reference point as described in Annex A to Recommendation P.64.

Transducers shall be stable and linear, and their physical design shall be such that they can be fitted in the handset chosen. A handset shall always contain both microphone and earphone capsules, irrespective of whether either is inactive during tests. The weight of a handset, so equipped, shall not exceed 350 g.

4 Subdivision of the complete IRS and impedances at the interfaces

Figure 1/P.48 shows the composition of the complete IRS, subdivided as specified in § 2 above. The principal features of the separate parts are considered below.

4.1 *Sending part*

The sending part of the IRS is defined as the portion A-JS extending from the handset microphone A to the interface with the junction at JS. The sending part shall include such amplification and equalization as necessary to ensure that the requirements of §§ 5.1 and 7 below are satisfied.

The return loss of the impedance at JS, towards A, against $600\underline{/0^\circ}$ ohms, when the sending part is correctly set up and calibrated, shall be not less than 20 dB over a frequency range 200-4000 Hz, and not less than 15 dB over a frequency range 125-6300 Hz.

4.2 *Receiving part*

The receiving part of the IRS is defined as the portion JR-B extending from the interface with the junction at JR to the handset earphone at B. The receiving part shall include such amplification and equalization as necessary to ensure that the requirements of §§ 5.2 and 7 below are satisfied.

The return loss of the impedance at JR, towards B, against $600\underline{/0^\circ}$ ohms, when the receiving part is correctly set up and calibrated, shall be not less than 20 dB over a frequency range 200-4000 Hz, and not less than 15 dB over a frequency range 125-6300 Hz.

4.3 *Junction*

For loudness balance and sidetone tests, the junction of the IRS shall comprise means of introducing known values of attenuation between the sending and receiving parts, and shall consist of a calibrated 600 ohm attenuator having a maximum value of not less than 100 dB

(e.g. 10×10 dB + 10×1 dB + 10×0.1 dB)

and having a tolerance, when permanently fitted and wired in position in the equipment, of not more than ± 1% of the dial reading or 0.1 dB, whichever is numerically greater. Provision shall be made for the inclusion of additional circuit elements (e.g. attenuation/frequency distortion) in the junction. The circuit configuration of such additional elements shall be compatible both with that of the attenuator and the junction interfaces. The return loss of the junction against $600\underline{/0^\circ}$ ohms, both with and without any additional circuit elements, shall be not less than 20 dB over a frequency range 200-4000 Hz, and not less than 15 dB over a frequency range 125-6300 Hz. For these tests, the port other than that being measured shall be closed with $600\underline{/0^\circ}$ ohms.

5 Nominal sensitivities of sending and receiving parts

The absolute values given below are provisional and may require changes to some extent as a result of the study of Question 19/XII [2].

5.1 *Sending part*

The sending sensitivity, S_{mJ} is given in Table 1/P.48, column (2) (see [3]).

5.2 *Receiving part*

The receiving sensitivity, S_{Je}, on a CCITT/IEC measured artificial ear (see Recommendation P.64) is given in Table 1/P.48, column (3) (see [3]).

TABLE 1/P.48

Nominal sending sensitivities and receiving sensitivities of the IRS

(These values were adopted provisionally)

Frequency (Hz)	S_{mJ}	S_{Je}
	dB V/Pa	dB Pa/V
(1)	(2)	(3)
100	−45.8	−27.5
125	−36.1	−18.8
160	−25.6	−10.8
200	−19.2	−2.7
250	−14.3	2.7
300	−11.3	6.4
315	−10.8	7.2
400	−8.4	9.9
500	−6.9	11.3
600	−6.3	11.8
630	−6.1	11.9
800	−4.9	12.3
1000	−3.7	12.6
1250	−2.3	12.5
1600	−0.6	13.0
2000	0.3	13.1
2500	1.8	13.1
3000	1.5	12.5
3150	1.8	12.6
3500	−7.3	3.9
4000	−37.2	−31.6
5000	−52.2	−54.9
6300	−73.6	−67.5
8000	−90.0	−90.0

6 Stability

The stability should be maintained, under reasonable ranges of ambient temperature and humidity, at least during the period between routine recalibrations. (See also [1].)

7 Shapes and tolerances on sensitivities of sending and receiving parts

The shape of the sensitivity/frequency characteristics of the sending and receiving parts of the IRS shall lie within the limits of masks formed by Table 2/P.48 and plotted in Figures 2/P.48 and 3/P.48. The sending and receiving loudness ratings shall both be set to 0 ± 0.2 dB when calculated in accordance with the principles laid down in Recommendation P.79.

Note – One excursion above or one excursion below the limits is permitted provided that:

a) the excursion is no greater than 2 dB above the upper or below the lower limit;

b) the width of the excursion as it breaks the appropriate limit is no greater than 1/10th of the frequency at the maximum or minimum of the excursion.

TABLE 2/P.48

Coordinates of sending and receiving sensitivity limit curves

Limite curve	Frequency (Hz)	Sending sensitivity (dB with respect to an arbitrary level)	Frequency (Hz)	Receiving sensitivity (dB with respect to an arbitrary level)
Upper limit	100 200 400 3400 3600 6000	−41 −16 −6 +6 +4 −60	100 200 300 500 3400 3600 4500	−24 0 +9 +14 +16 +13 −40
Lower limit	Under 200 200 400 3000 3400 Over 3400	−∞ −21 −11 −1 −4 −∞	Under 200 200 300 500 3200 3400 Over 3400	−∞ −20 +4 +9 +10 +4 −∞

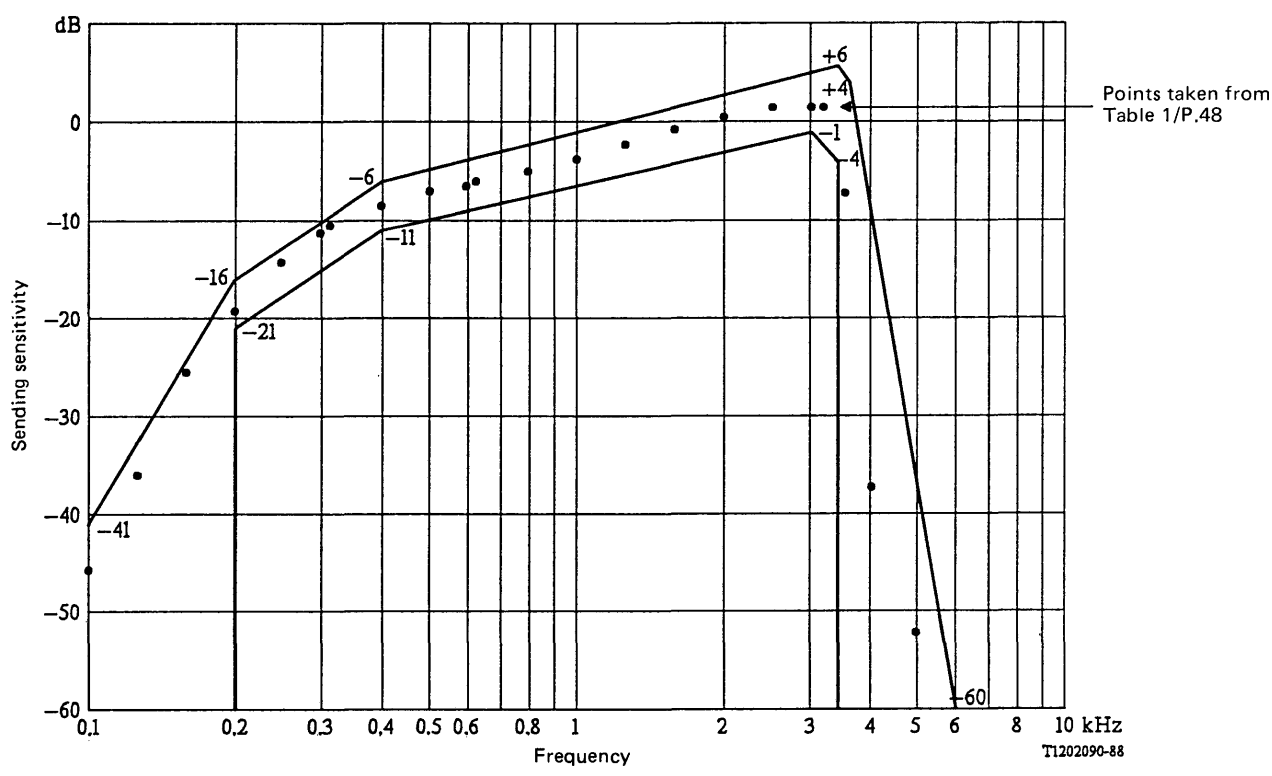


FIGURE 2/P.48

Suggested IRS sending mask

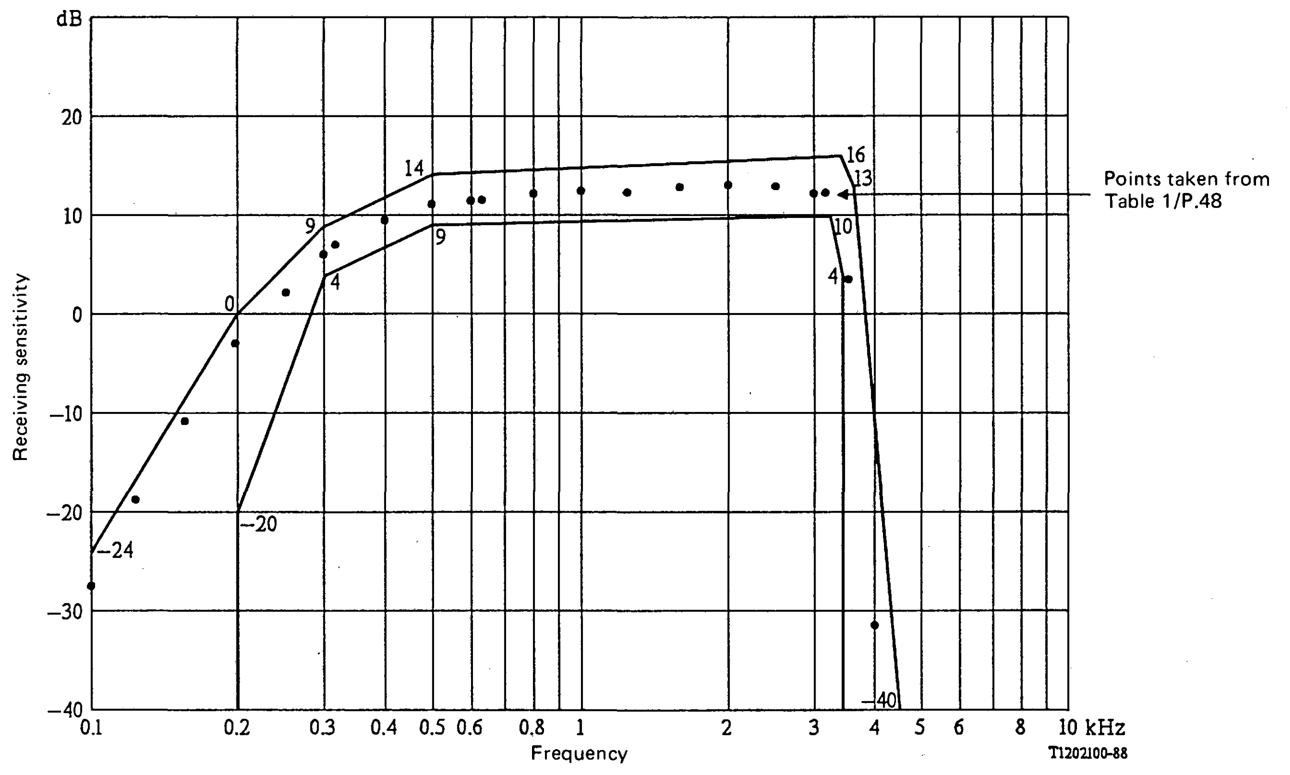


FIGURE 3/P.48

Suggested IRS receiving mask

8 Noise limits

It is important that the noise level in the system be well controlled. See [4].

9 Nonlinear distortion

In order to ensure that nonlinear distortion will be negligible with the vocal levels normally used for loudness rating, requirements in respect of distortion shall be met.

10 Complete specifications

Certain secondary characteristics of an IRS may be included in Administrations' specifications. Particularly, special care must be given to adjustable components, stability and tolerances, crosstalk, installation and maintenance operations, etc. Reference [1] gives some guidance on these points.

References

[1] *Precautions to be taken for correct installation and maintenance of an IRS,* Orange Book, Vol. V, Supplement No. 1, ITU, Geneva, 1977.

[2] CCITT – Question 19/XII, Contribution COM XII-No. 1, Study Period 1985-1988, ITU, Geneva, 1985.

[3] *Precautions to be taken for correct installation and maintenance of an IRS,* Orange Book, Vol. V, Supplement No. 1, § 9.2, ITU, Geneva, 1977.

[4] *Ibid.,* § 5.

SECTION 4

OBJECTIVE MEASURING APPARATUS

Recommendation P.50

ARTIFICIAL VOICES

(Melbourne, 1988)

The CCITT,

considering

(a) that it is highly desirable to perform objective telephonometric measurements by means of a mathematically defined signal reproducing the characteristics of human speech;

(b) that the standardization of such a signal is a subject for general study by the CCITT,

recommends[1)]

the use of the artificial voice described in this Recommendation.

Note 1 – For objective loudness rating measurements, less sophisticated signals such as pink noise or spectrum-shaped Gaussian noise can be used instead of the artificial voice.

Note 2 – The artificial voice here recommended has not yet been exhaustively tested in all possible applications; further studies being carried out within Question 14/XII.

1 Introduction

The signal here described reproduces the characteristics of human speech for the purposes of characterizing linear and nonlinear telecommunication systems and devices, which are intended for the transduction or transmission of speech. It is known that for some purposes, such as objective loudness rating measurements, more simple signals can be used as well. Examples of such signals are pink noise or spectrum-shaped Gaussian noise, which nevertheless cannot be referred to as "artificial voice" for the purpose of this Recommendation.

The artificial voice is a signal that is mathematically defined and that reproduces the time and spectral characteristics of speech which significantly affect the performances of telecommunication systems [1]. Two kinds of artificial voice are defined, reproducing respectively the spectral characteristics of female and male speech.

The following time and spectral characteristics of real speech are reproduced by the artificial voice:

a) long-term average spectrum,

b) short-term spectrum,

c) instantaneous amplitude distribution,

d) voiced and unvoiced structure of speech waveform,

e) syllabic envelope.

1) The specifications given here are subject to future enhancement and therefore should be regarded as provisional.

2 Scope, purpose and definition

2.1 *Scope and purpose*

The artificial voice is aimed at reproducing the characteristics of real speech over the bandwidth 100 Hz – 8 kHz. It can be utilized for characterizing many devices, e.g. carbon microphones, loudspeaking telephone sets, nonlinear coders, echo controlling devices, syllabic compandors, nonlinear systems in general.

The use of the artificial voice instead of real speech has the advantage of both being more easily generated and having a smaller variability than samples of real voice.

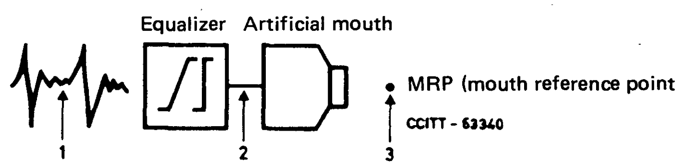
Of course, when a particular system is tested, the characteristics of the transmission path preceding it are to be considered. The actual test signal has then to be produced as the convolution between the artificial voice and the path response.

2.2 *Definition*

The artificial voice is a signal, mathematically defined, which reproduces all human speech characteristics, relevant to the characterization of linear and nonlinear telecommunication systems. It is intended to give a satisfactory correlation between objective measurements and real speech tests.

3 Terminology

The artificial voice can be produced both as an electric or as an acoustic signal, according to the system or device under test (e.g. communication channels, coders, microphones). The following definitions apply with reference to Figure 1/P.50.



1: Electrical artificial voice
2: Artificial mouth excitation signal
3: Acoustic artificial voice

FIGURE 1/P.50

3.1 *electrical artificial voice*

The artificial voice produced as an electrical signal, used for testing transmission channels or other electric devices.

3.2 *artificial mouth excitation signal*

A signal applied to the artificial mouth in order to produce the acoustic artificial voice. It is obtained by equalizing the electrical artificial voice for compensating the sensitivity/frequency characteristic of the mouth.

Note 1 – The equalization depends on the particular artificial mouth employed and can be accomplished electrically or mathematically within the signal generation process.

3.3 **acoustic artificial voice**

It is the acoustic signal at the MRP (Mouth Reference Point) of the artificial mouth and has to comply with the same time and spectral requirements of the electrical artificial voice.

4 Characteristics

4.1 *Long-term average spectrum*

The third octave filtered long-term average spectrum of the artificial voice is given in Figure 2/P.50 and Table 1/P.50, normalized for a wideband sound pressure level of −4.7 dBPa. The table is calculated from the theoretical equation reported in [2].

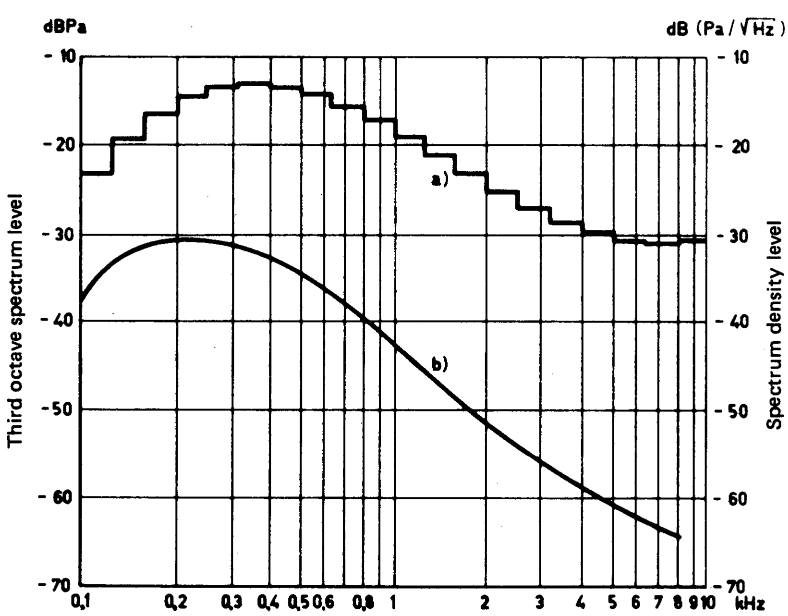
Note – The values of the long-term spectrum of the artificial voice at the MRP can be derived from the equation:

$$S(f) = -376.44 + 465.439(\log_{10} f) - 157.745(\log_{10} f)^2 + 16.7124(\log_{10} f)^3 \qquad (1\text{-}1)$$

where $S(f)$ is the spectrum density in dB relative to 1 pW/m^2 sound intensity per Hertz at the frequency f. The definition frequency range is from 100 Hz to 8 kHz.

The curve of the spectrum is shown in Figure 2/P.50. The values of $S(f)$ at 1/3 octave ISO frequencies are given in the fourth column of Table 1/P.50. The tolerances are given in the fifth column of Table 1/P.50. The tolerances below 200 Hz apply onto to the male artificial voice.

The total sound pressure level of the spectrum defined in Equation (1-1) is −4.7 dBPa. However, this spectrum is also applicable for the levels from −19.7 to +10.3 dPBa. In other words, the first term of Equation (1-1) may range from −391.44 to −361.44.



a) Third octave spectrum [Column (3), Table 1/P.50].

b) Spectrum density [Column (3)-(2), Table 1/P.50].

FIGURE 2/P.50

Long-term spectrum of artificial voice

TABLE 1/P.50

Long-term spectrum of the artificial voice

1/3 octave center frequency (Hz) (1)	Bandwidth correction factor $10 \log_{10} \Delta f$ (dB) (2)	Sound pressure level (third octave) (dBPa) (3)	Spectrum density (dB) (3) − (2)	Tolerance (dB)
100	13.6	−23.1	−36.7	−
125	14.6	−19.2	−33.8	+3, −6 [a)]
160	15.6	−16.4	−32	+3, −6 [a)]
200	16.6	−14.4	−31	+3, −6
250	17.6	−13.4	−31	±3.0
315	18.6	−13.0	−31.6	±3.0
400	19.6	−13.3	−32.9	±3.0
500	20.6	−14.1	−34.7	±3.0
630	21.6	−15.4	−37	±3.0
800	22.6	−17.0	−39.6	±3.0
1000	23.6	−18.9	−42.5	±3.0
1250	24.6	−21.0	−45.6	±3.0
1600	25.6	−23.0	−48.6	±3.0
2000	26.6	−25.1	−51.7	±3.0
2500	27.6	−26.9	−54.5	±3.0
3150	28.6	−28.6	−57.2	±3.0
4000	29.6	−29.8	−59.4	±6.0
5000	30.6	−30.6	−61.2	±6.0
6300	31.6	−30.9	−62.5	±6.0
8000	32.6	−30.5	−63.1	−

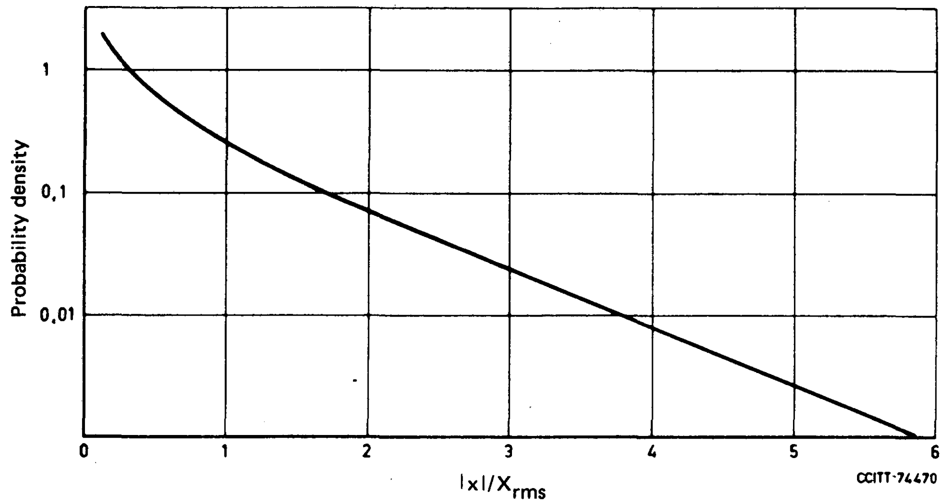
a) The given tolerances apply to the long-term spectrum of male speech and must also be complied with by speech shaped noises. However, they do not apply to the female speech spectrum, whose energy content in this frequency range is virtually negligible.

4.2 *Short-term spectrum*

The short-term spectrum characteristics of the male and female artificial voices are described in Annex A.

4.3 *Instantaneous amplitude distribution*

The probability density distribution of the instantaneous amplitude of the artificial voice is shown in Figure 3/P.50 [3].



|x| : absolute value of the instantaneous amplitude
X_{rms}: root mean square of the signal

FIGURE 3/P.50

Instantaneous amplitude distribution

4.4 *Segmental power level distribution*

The segmental power level distribution of the artificial voice, measured on time windows of 16 ms, is shown in Figure 4/P.50. The upper and lower tolerance limits are reported as well.

Note – The upper tolerance limit represents the typical segmental power level distribution of normal conversation, while the lower limit represents continuous speech (telephonometric phrases) [4], [5].

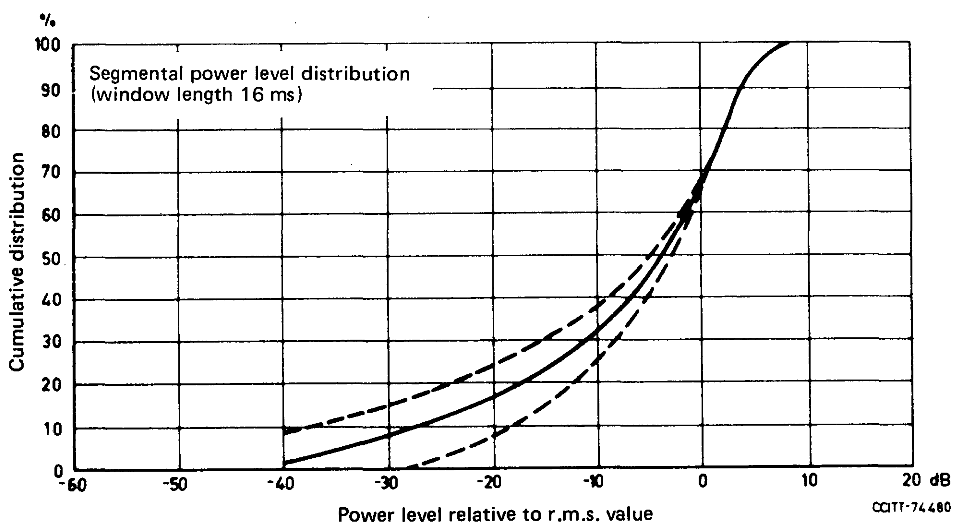


FIGURE 4/P.50

Segmental power level distribution

4.5 *Spectrum of the modulation envelope*

The spectrum of the modulation envelope waveform is shown in Figure 5/P.50 and should be reproduced with a tolerance of ± 5 dB on the whole frequency range.

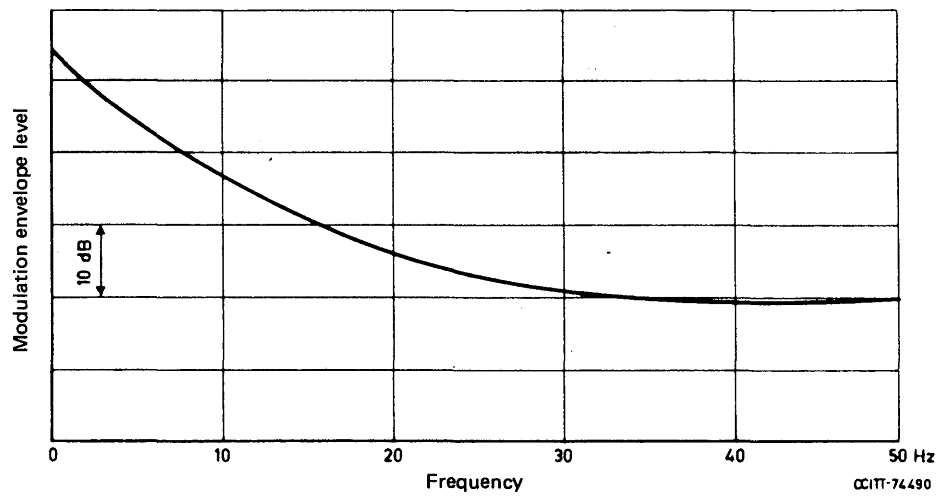


FIGURE 5/P.50

Spectrum of modulation envelope

4.6 *Time convergence*

The artificial voice must exhibit characteristics as close as possible to real speech. Particularly, it should be possible to obtain the long-term spectrum and amplitude distribution characteristics in 10 s.

5 Generation method

Figure 6/P.50 shows a block diagram of the generation process of the artificial voice. It is generated by applying two different types of excitation source signals, a glottal excitation signal and a random noise, to a time-variant spectrum shaping filter. The artificial voice generated by the glottal excitation signal and by the random noise corresponds respectively to voiced and unvoiced sounds. The frequency response of the spectrum shaping filter simulates the transmission characteristics of the vocal tract.

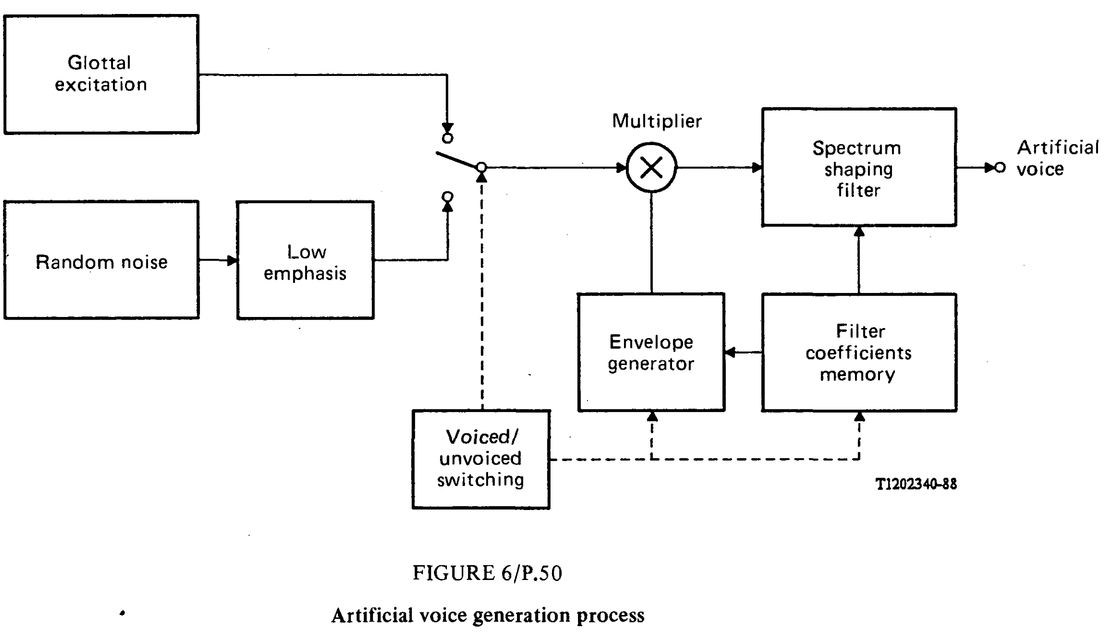


FIGURE 6/P.50

Artificial voice generation process

5.1 *Excitation source signal*

The artifical voice is obtained by randomly alternating four basic unit elements, each containing voiced and unvoiced segments. While one unit element starts with an unvoiced sound, followed by a voiced one, the other three elements start with a voiced sound, followed by an unvoiced one and end with a voiced sound again (see also Figure 9/P.50). The ratio of the unvoiced sound duration T_{uv} to the total duration of voiced segments T_v for each unit element is 0.25. The duration $T = T_{uv} + T_v$ of unit elements varies according to the following equation:

$$T = -3.486 \ (\log_{10} r)$$

where r denotes a uniformly distributed random number ($0.371 \leqslant r \leqslant 0.609$).

The time lengths of the voiced and unvoiced sounds of the four unit elements are as follows:

Element a: Unvoiced (T_{uv}) ; Voiced (T_v)

Element b: Voiced ($T_v/4$) + Unvoiced (T_{uv}) + Voiced ($3T_v/4$)

Element c: Voiced ($T_v/2$) + Unvoiced (T_{uv}) ; Voiced ($T_v/2$)

Element d: Voiced ($3T_v/4$) + Unvoiced (T_{uv}) + Voiced ($T_v/4$)

Unit elements shall be randomly iterated for at least 10 s in order to comply with the artificial voice characteristics as specified in § 4.

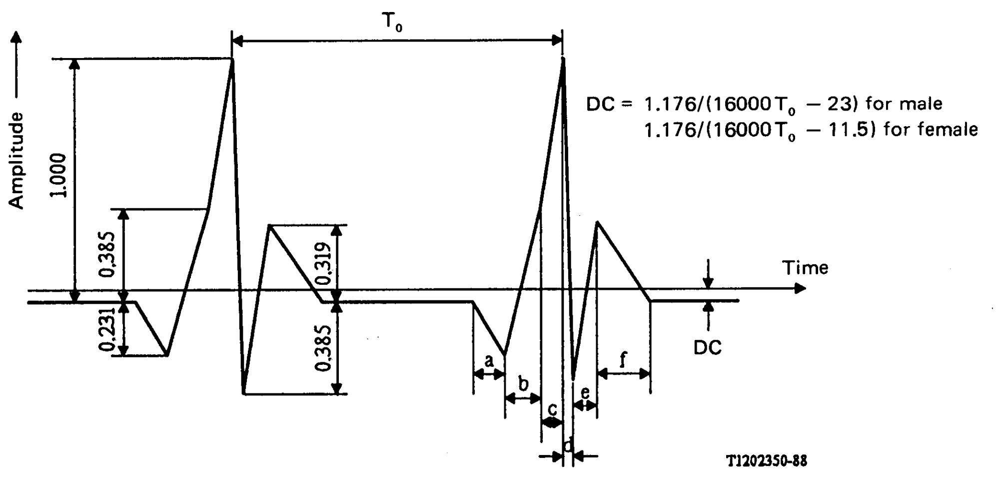
5.2 *Glottal excitation*

The glottal excitation signal is a periodic waveform as shown in Figure 7/P.50. The pitch frequency ($1/T_0$ in Figure 7/P.50) varies according to the variation pattern shown in Figure 8/P.50 during the period T_v. The starting value of the pitch frequency (F_s in Figure 8/P.50) is determined according to the following relationships:

$F_s = F_c - 31.82 \ T_v + 39.4 \ R$ for the male artificial voice

$F_s = F_c - 51.85 \ T_v + 64.2 \ R$ for the female artificial voice

where F_c and R respectively denote the center frequency and a uniformly distributed random variable ($-1 < R < 1$). F_c is 128 Hz for the male artificial voice and 215 Hz for the female artificial voice. In the trapezoid of the pitch frequency variation pattern, the area of the trapezoid above F_c should be equal to that below F_c (shaded in Figure 8/P.50). For the elements b), c) and d) in Figure 7/P.50 the pitch frequency variation pattern applies to the combination of the two voiced parts, irrespectively of where the unvoiced segment is inserted.



a = 0.375 ms
b = 0.5 ms
c = e = 0.3125 ms
d = 0.1875 ms
f = 0.625 ms

FIGURE 7/P.50

Glottal excitation signal

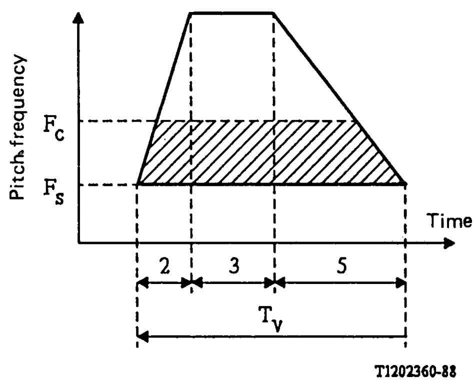


FIGURE 8/P.50

Pitch frequency variation pattern

5.3 *Unvoiced sounds*

The transfer function of the low-pass filter located after the random noise generator (low emphasis) is $1/(1 - z_{-1})$, where z^{-1} denotes the unit delay.

5.4 *Power envelope*

The power envelope of each unit element of the excitation source signal is so controlled that the short-term segmental power (evaluated over 2 ms intervals) of the artificial voice varies according to the patterns shown in a) to d) of Figure 9/P.50. This is obtained by utilizing the following relationship providing input and output signals of the spectrum shaping filter:

$$P_{in} = P_{out} \prod_{i=1}^{12} (1 - k_i^2)$$

where:

P_{in} is the input power to the spectrum shaping filter

P_{out} is the output power from the spectrum shaping filter

k_i is the ith coefficient of the spectrum shaping filter.

The rising, stationary and decay times of each trapezoid of a) to d) of Figure 9/P.50 shall be mutually related by the same proportionality coefficients (2 : 3 : 5) of the pitch frequency variation pattern shown in Figure 8/P.50. For each unit element, the average power of unvoiced sounds (P_{uv}) shall be 17.5 dB less than the average power of voiced sounds (P_v).

5.5 *Spectrum shaping filter*

The spectrum shaping filter has a 12th order lattice structure as shown in Figure 10/P.50. Sixteen groups, each of 12 filtering coefficients ($k_1 - k_{12}$), are defined; thirteen groups shall be used for generating the voiced part, while three groups shall be used for generating the unvoiced part. These coefficients are listed in Table 2/P.50 both for male and female artificial voices.

The twelve filter coefficients shall be updated every 60 ms while generating the signal. More precisely, during each 60 ms period the actual filtering coefficients must be adjourned every 2 ms, by linearly interpolating between the two sets of values adopted for subsequent 60 ms intervals. In the voiced sound part, each of 13 groups of coefficients shall be chosen at random once every 780 ms (= 60 ms × 13), and in the unvoiced sound part each of 3 groups of coefficients shall be chosen at random once every 180 ms (= 60 ms × 3).

Note – The described implementation of the shaping filter should be considered as an example and is not an integral part of this Recommendation. Any other implementation providing the same transfer function can be alternatively used.

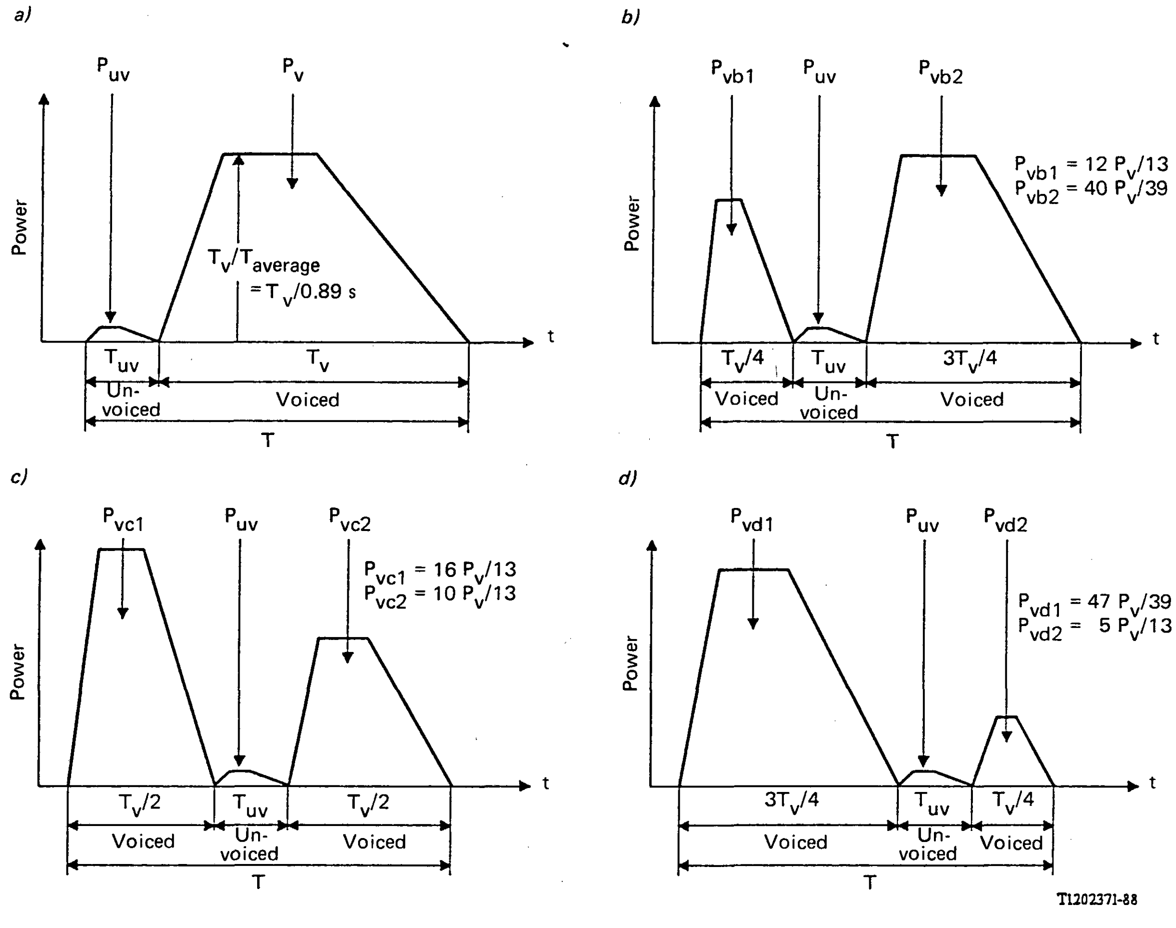


FIGURE 9/P.50

Short-term power variation patterns of the four unit elements used to generate the artificial voice

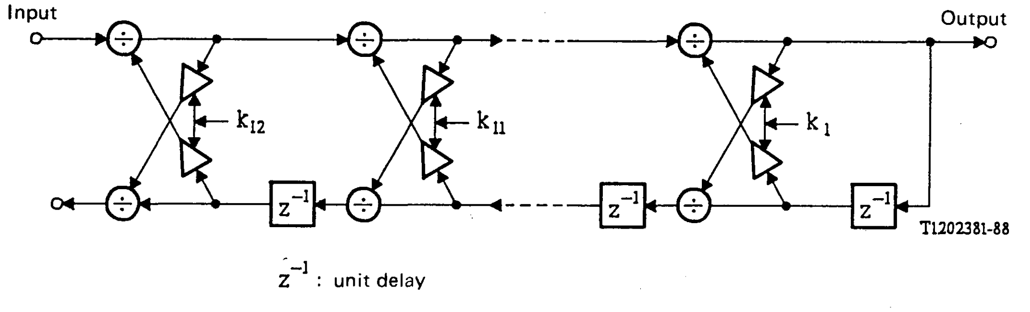


FIGURE 10/P.50

Spectrum shaping filter

TABLE 2/P.50

Coefficients k_i

a) k_i for male artificial voice

		k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}	k_{11}	k_{12}
Unvoiced	1	-0.471	-0.108	0.024	-0.048	0.140	0.036	0.054	0.004	0.123	0.044	0.099	-0.003
	2	-0.284	-0.468	0.030	0.090	0.124	-0.020	0.087	0.067	0.131	0.011	0.076	-0.024
	3	-0.025	-0.496	-0.176	0.162	0.236	-0.012	0.068	0.001	0.096	0.029	0.086	-0.018
Voiced	1	0.974	0.219	0.025	-0.123	-0.132	-0.203	-0.103	-0.174	-0.079	-0.153	-0.010	-0.061
	2	0,629	-0.152	-0.138	-0.142	-0.118	-0.135	0.147	0.019	0.077	-0.040	0.029	-0.007
	3	0.599	-0.119	0.067	0.051	0.103	0.023	0.106	0.036	-0.006	-0.133	-0.052	-0.094
	4	0.164	-0.364	-0.248	-0.076	0.168	0.072	0.103	0.045	0.112	0.010	0.048	-0.034
	5	0.842	0.022	0.171	0.173	0.067	-0.057	0.089	-0.045	-0.039	-0.134	-0.034	-0.122
	6	0.933	-0.537	-0.137	-0.161	-0.216	-0.139	0.115	-0.042	0.027	-0.163	0.102	-0.107
	7	0.937	-0.413	0.132	-0.059	-0.103	-0.134	0.047	-0.115	-0.105	-0.097	0.039	-0.108
	8	0.965	-0.034	0.032	0.001	-0.107	-0.189	-0.057	-0.175	-0.109	-0.163	-0.003	-0.055
	9	0.870	-0.476	-0.016	-0.136	-0.125	-0.107	0.091	-0.008	0.021	-0.128	0.042	-0.069
	10	0.686	-0.030	0.178	0.197	0.155	-0.026	0.078	0.004	-0.001	-0.128	-0.004	-0.102
	11	0.963	-0.232	0.086	-0.018	-0.147	-0.192	-0.040	-0.179	-0.144	-0.133	0.042	-0.042
	12	0.930	-0.461	0.071	-0.144	-0.122	-0.096	0.034	-0.066	-0.021	-0.171	0.067	-0.091
	13	0.949	-0.334	0.143	-0.040	-0.112	-0.161	0.010	-0.156	-0.123	-0.119	0.049	-0.070

b) k_i for female artificial voice

		k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}	k_{11}	k_{12}
Unvoiced	1	-0.488	-0.388	0.145	0.053	0.122	0.027	0.135	0.035	0.080	0.017	0.068	-0.028
	2	-0.093	-0.444	-0.102	0.121	0.154	0.009	0.102	-0.031	0.084	0.019	0.101	-0.020
	3	-0.709	-0.179	0.134	0.007	0.142	0.027	0.099	0.000	0.115	0.007	0.075	-0.037
Voiced	1	0.355	-0.247	-0.092	-0.043	0.032	0.046	0.113	-0.023	0.071	-0.030	-0.000	-0.116
	2	0.976	0.150	-0.062	-0.187	-0.172	-0.200	-0.122	-0.207	-0.054	-0.127	0.012	-0.111
	3	0.737	-0.324	-0.175	-0.197	-0.153	0.023	0.110	-0.018	0.040	-0.062	0.034	-0.091
	4	0.598	0.234	0.126	0.011	-0.005	-0.026	0.131	0.032	0.073	-0.063	0.011	-0.088
	5	0.808	0.118	0.262	0.139	0.063	-0.024	0.001	-0.184	-0.056	-0.100	0.014	-0.115
	6	0.914	-0.500	-0.051	-0.115	-0.211	-0.012	-0.077	-0.179	0.064	-0.102	0.037	-0.092
	7	0.933	-0.359	0.089	-0.107	-0.178	-0.050	-0.137	-0.206	0.046	-0.088	-0.004	-0.074
	8	0.966	-0.023	0.044	-0.105	-0.178	-0.195	-0.150	-0.233	-0.045	-0.092	0.029	-0.097
	9	0.870	-0.469	-0.244	-0.107	-0.140	-0.037	0.084	-0.131	0.021	-0.066	-0.003	-0.091
	10	0.673	-0.292	0.392	0.158	0.143	0.160	0.019	-0.281	-0.105	-0.195	-0.156	-0.185
	11	0.962	-0.191	0.030	-0.089	-0.207	-0.133	-0.141	-0.263	0.007	-0.054	0.014	-0.074
	12	0.879	-0.340	0.046	-0.049	-0.071	-0.024	-0.039	-0.188	0.017	-0.078	-0.014	-0.117
	13	0.941	-0.258	0.122	-0.073	-0.163	-0.089	-0.151	-0.250	0.025	-0.062	-0.006	-0.093

ANNEX A

(to Recommendation P.50)

Short-term spectrum characteristics of the artificial voice

The artificial voice is generated by randomly selecting each of sixteen short-term spectrum patterns once ever 960 ms (= 60 ms × 16 patterns). The spectrum density of each pattern is provided by Equation (A-1) and Table A-1/P.50, and the short-term spectrum of the signal during the 60 ms interval occurring between any two subsequent pattern selections varies smoothly from one pattern to the next.

Note – The spectrum patterns in Equation (A-10) and Table A-1/P.50 are expressed in power normalized form.

$$\text{Spectrum density } S_i(f) = \frac{1}{A_{io} + 2 \sum_{j=1}^{12} A_{ij} [\cos(2\pi jf)]}, \quad i = 1, 2, \ldots 16 \qquad \text{(A-1)}$$

TABLE A-1/P.50

Coefficients A_{ij}

a) A_{ij} for male artificial voice

i \ j	0	1	2	3	4	5	6	7	8	9	10	11	12
1	2.09230	−1.33222	1.32175	−1.14200	0.99352	−0.94634	0.72684	−0.63263	0.41196	−0.42858	0.22070	−0.19746	0.10900
2	9.34810	−8.55934	7.35732	−6.35320	5.33999	−4.47238	3.62417	−2.85246	2.12260	−1.49424	0.93988	−0.44998	0.12400
3	11.69068	−10.91138	9.46588	−8.11729	6.94160	−5.90977	4.95137	−3.89587	2.88750	−1.97671	1.14892	−0.50255	0.12100
4	12.56830	−11.81209	10.36030	−8.82879	7.37947	−6.01017	4.66740	−3.46913	2.42182	−1.60880	0.91652	−0.39648	0.12000
5	6.83438	−6.18275	5.59089	−4.71866	4.06004	−3.44767	2.65380	−2.12140	1.50334	−1.07904	0.64553	−0.31816	0.11500
6	12.37251	−11.52358	9.89962	−8.31774	6.99062	−5.86272	4.69809	−3.56806	2.53340	−1.70522	0.99232	−0.45403	0.13400
7	21.07637	−19.62125	16.56781	−13.67518	11.41379	−9.61940	7.93529	−6.32841	4.92443	−3.53539	2.09095	−0.86543	0.18100
8	30.77371	−29.17365	25.52254	−21.51978	17.80583	−14.30488	10.87190	−7.71572	5.14643	−3.20113	1.72149	−0.68054	0.14400
9	4.18618	−3.36611	3.36793	−2.92133	2.38452	−2.06047	1.57550	−1.34240	0.84994	−0.70462	0.38685	−0.21857	0.12100
10	14.12359	−13.14611	11.25804	−9.47510	7.97588	−6.70717	5.44803	−4.23843	3.10807	−2.12879	1.25096	−0.53230	0.12600
11	26.36971	−24.95984	21.80496	−18.41045	15.30642	−12.49415	9.84879	−7.40287	5.29262	−3.43906	1.84980	−0.71546	0.14800
12	11.50808	−10.74609	9.34328	−7.91953	6.66959	−5.54500	4.34328	−3.27036	2.33714	−1.61333	0.96597	−0.44666	0.13500
13	5.32020	−4.61998	4.29145	−3.62118	3.01310	−2.67071	2.13992	−1.72147	1.22163	−0.93163	0.53317	−0.28989	0.11900
14	20.61945	−19.39682	16.80034	−14.14817	11.84307	−9.78712	7.73534	−5.77921	4.06200	−2.66324	1.49831	−0.59887	0.12600
15	30.02641	−28.42244	24.75314	−20.70178	16.98199	−13.72247	10.81050	−8.20966	5.94148	−3.90501	2.11507	−0.81306	0.16400
16	27.62370	−26.17896	22.93678	−19.42253	16.18997	−13.17171	10.19859	−7.42299	5.07437	−3.21481	1.73980	−0.67818	0.14000

b) A_{ij} for female artificial voice

i \ j	0	1	2	3	4	5	6	7	8	9	10	11	12
1	8.92953	−8.28905	7.23150	−6.06571	5.06663	−4.16883	3.34820	−2.64174	1.91152	−1.27122	0.74358	−0.35347	0.13100
2	9.11050	−8.29868	7.05018	−6.03862	5.02156	−4.15784	3.37442	−2.70084	2.04257	−1.41928	0.90339	−0.47240	0.14100
3	13.69058	−12.75539	10.87390	−8.98976	7.28838	−5.87662	4.78110	−3.72852	2.71831	−1.81828	1.06757	−0.48907	0.12900
4	6.95118	−6.29134	5.35757	−4.64910	3.87250	−3.23132	2.69856	−2.12367	1.58326	−1.13885	0.70609	−0.37778	0.12800
5	2.73454	−1.80664	1.95283	−1.79464	1.43897	−1.31656	0.93268	−0.87398	0.53694	−0.47562	0.24159	−0.15438	0.11800
6	10.82358	−10.07808	8.78565	−7.43643	6.14765	−5.10041	4.10027	−3.23241	2.34220	−1.54676	0.91918	−0.45059	0.14200
7	21.58481	−19.92676	16.21532	−12.43566	9.61057	−7.93982	6.97363	−6.00866	4.68271	−3.12797	1.70804	−0.67499	0.14200
8	23.73912	−22.20897	18.74416	−15.03715	11.99248	−9.85513	8.27112	−6.72826	4.94335	−3.10450	1.60004	−0.61090	0.12900
9	4.97162	−4.27705	4.01380	−3.38500	2.78457	−2.45010	1.98057	−1.63020	1.18104	−0.80108	0.51528	−0.29138	0.12500
10	13.37598	−12.45509	10.72295	−8.97928	7.35893	−6.05438	4.88819	−3.86108	2.85164	−1.88876	1.11490	−0.52260	0.13800
11	16.48817	−15.22287	12.62125	−10.23900	8.46966	−7.23692	6.24648	−5.09752	3.77465	−2.46950	1.37404	−0.57453	0.13200
12	18.22041	−17.17540	15.09489	−12.55171	10.24976	−8.45903	6.71874	−5.19063	3.52021	−2.10167	1.08066	−0.41880	0.14300
13	1.32602	−0.31718	0.44277	−0.47070	0.32935	−0.41555	0.25775	−0.32079	0.13791	−0.23640	0.10103	−0.10136	0.11500
14	16.90640	−15.73723	13.30151	−10.82887	8.78690	−7.34521	6.21516	−5.11100	3.80281	−2.43990	1.33506	−0.55971	0.12900
15	21.73895	−20.42432	17.51117	14.44152	11.79131	−9.66735	7.90433	−6.19508	4.41275	−2.75545	1.46525	−0.59916	0.14000
16	21.04832	−19.72714	16.81197	−13.70183	11.07189	−9.12707	7.57941	−6.08064	4.40471	−2.74320	1.43897	−0.58079	0.13300

References

[1] CCITT – Contribution COM XII-No. 76, Study Period 1981-1984

[2] CCITT – Contribution COM XII-No. 108, Study Period 1981-1984

[3] CCITT – Contribution COM XII-No. 11, Study Period 1981-1984

[4] CCITT – Contribution COM XII-No. 150, Study Period 1981-1984

[5] CCITT – Contribution COM XII-No. 132, Study Period 1981-1984

Recommendation P.51

ARTIFICIAL EAR AND ARTIFICIAL MOUTH

(amended at Mar del Plata, 1968, Geneva, 1972, 1976, 1980, Malaga-Torremolinos, 1984 and Melbourne, 1988)

The CCITT,

considering

(a) that it is highly desirable to design an apparatus for telephonometric measurements such that in the future all of these measurements may be made with this apparatus, without having recourse to the human mouth and ear;

(b) that the standardization of the artificial ear and mouth used in the construction of such apparatus is a subject for general study by the CCITT,

recommends

(1) the use of the artificial ears described in § 1 of this Recommendation;

(2) the use of the artificial mouth described in § 2 of this Recommendation.

Note – Administrations may, if they wish, use devices which they have been able to construct for large-scale testing of telephone apparatus supplied by manufacturers, provided that the results obtained with these devices are in satisfactory agreement with results obtained by real voice-ear methods.

1 Artificial ears

Three types of artificial ears are defined:

1) a wideband type for audiometricand telephonometric measurements,
2) a special type for measuring insert earphones,
3) a type which faithfully reproduces the characteristics of the average human ear, for use in the laboratory.

Type 1 is covered by IEC Recommendation 318 [1], the second IEC Recommendation 711 [2] and the third is the object of further study in the IEC.

It is recommended that the artificial ear conforming to IEC 318 [1] should be used for measurements on supra-aural earphones, e.g. handsets, and that the insert ear simulator conforming to IEC 711 [2] should be used for measurements on insert earphones, e.g. some headsets.

Note 1 – For the calibration of NOSFER earphones with rubber earpads (types 4026A and DR 701) the method detailed in Annex B to Recommendation P.42 should be used.

Note 2 – The sound pressure measured by the IEC 711 artificial ear is referred to the eardrum. The correction function given in Table 1/P.51 shall be used for converting data to the ear reference point (ERP), where loudness rating algorithms (Recommendation P.79) are based. The corrections apply to free field open-ear conditions and to partially or totally occluded conditions as well.

TABLE 1/P.51

Frequency (Hz)	S_{DE} (dB)
100	0.0
125	0.0
160	0.0
200	0.0
250	0.0
315	−0.2
400	−0.5
500	−1.1
630	−1.0
800	−1.8
1000	−2.0
1250	−2.5
1600	−4.1
2000	−7.2
2500	−10.6
3150	−10.4
4000	−6.0
5000	−2.1

S_{DE} is the transfer function eardrum to ERP:

$S_{DE} = 20 \log \frac{P_E}{P_D}$ (dB), where

P_E sound pressure at the ERP

P_D sound pressure at the eardrum.

2 Artificial mouth

2.1 *Introduction*

The artificial mouth is a device that accurately reproduces the acoustic field generated by the human mouth in the near field. It is used for measuring objectively the sending characteristics of handset-equipped telephone sets as specified in Recommendation P.64. It may also be used for measuring the sending characteristics of loudspeaking telephones at distances up to 0.5 m from the lip plane, but the accuracy with which it reproduces the sound field of the human mouth is slightly reduced.

2.2 *Definitions*

2.2.1 **lip ring**

Circular ring of thin rigid rod, having a diameter of 25 mm and less than 2 mm thick. It shall be constructed of non-magnetic material and be solidly fixed to the case of the artificial mouth. The lip ring defines both the reference axis of the mouth and the mouth reference point.

Note – The provision of the lip ring for locating the lip planes and the reference axis is not mandatory. However, when not provided, adequate markings or other suitable geometric reference shall be alternatively available.

2.2.2 **lip plane**

Outer plane of the lip ring.

2.2.3 **reference axis**

The line perpendicular to the lip plane containing the center of the lip ring.

2.2.4 **vertical plane**

A plane containing the reference axis that divides the mouth into symmetrical halves. It shall be vertically oriented in order to reproduce the acoustic field generated by a person in the upright position.

2.2.5 **horizontal plane**

The plane containing the reference axis, perpendicular to the vertical plane. It shall be horizontally oriented in order to reproduce the acoustic field generated by a person in the upright position.

2.2.6 **mouth reference point (MRP)**

The point on the reference axis, 25 mm in front of the lip plane.

2.2.7 **normalized free-field response (at a given point)**

Difference between the third-octave spectrum level of the signal delivered by the artificial mouth at a given point in the free field and the third-octave spectrum level of the signal delivered simultaneously at the MRP. The characteristic is measured by feeding the artificial voice (see Recommendtion P.50) a speech-shaped random noise or a pink noise.

2.2.8. **reference obstacle**

Disc constructed of hard, stable and on-megnetic material, such as brass, having a diameter of 63 mm and 5 mm thick. In order to measure the normalized obstacle diffraction, it shall be fitted with a ¼″ pressure microphone, mounted at the centre with the diaphragm flush on the disc surface.

2.2.9 **normalized obstacle diffraction**

Difference between the third-octave spectrum level of the acoustic pressure delivered by the artificial mouth at the surface of the reference obstacle and the third-octave spectrum level of the pressure simultaneously delivered at the point on the reference axis, 500 mm in front of the lip plane. The characteristic is defined for positions of the reference obstacle in front of the artificial mouth, with the disc axis coinciding with the reference axis, and is measured by feeding the artificial mouth with a complex signal such as the artificial voice, a speech shaped random noise or a pink noise.

2.3 *Acoustic characteristics of the artificial mouth*

2.3.1 *Normalized free-field response*

The normalized free-field response is specified at seventeen points: ten in the near field and seven in the far field. Near-field points are listed in Table 2/P.51, while far-field points are listed in Table 3/P.51.

Table 4/P.51 provides the normalized free-field response of the artificial mouth, together with tolerances, for the bandwidth between 100 Hz and 8 kHz. The requirements at each point not lying in the vertical plan shall also be met by the corresponding point in the symmetrical half-space.

The characteristic shall be checked by using appropriate microphones, as specified in Table 5/P.51. Pressure microphones shall be oriented with their axes perpendicular to the sound direction, while free-field microphones shall be oriented with their axes parallel to the direction of sound.

Note – If a compressor microphone is used with the mouth, it (or an equivalent dummy) shall be left in place while checking the normalized free-field response.

TABLE 2/P.51

Coordinates of points in the near field

Measurement point	On-axis displacement from the lip plane (mm)	Off-axis, perpendicular displacement (mm)
1	12.5	0
2	50	0
3	100	0
4	140	0
5	0	20 horizontal
6	0	40 horizontal
7	25	20 horizontal
8	25	40 horizontal
9	25	20 vertical (downwards)
10	25	40 vertical

TABLE 3/P.51

Coordinates of points in the far field

Measurement point	Distance from the lip plane (mm)	Azimuth angle (horizontal) (degree)	Elevation angle (vertical) (degree)
11	500	0	0
12	500	0	+15
13	500	0	+30
14	500	0	−15
15	500	0	−30
16	500	15	0
17	500	30	0

TABLE 4a/P.51

Normalized free field response at points on axis in the near field

Frequency	Measurement point				
	1	2	3	4	Tolerance
(Hz)	(dB)	(dB)	(dB)	(dB)	(dB)
100	4.2	−5.0	−11.0	−13.6	±1.5
125	4.2	−5.0	−10.9	−13.6	±1.5
160	4.2	−5.0	−10.7	−13.6	±1.5
200	4.0	−5.0	−10.7	−13.3	±1.5
250	4.0	−5.0	−10.6	−13.2	±1.5
315	4.0	−5.0	−10.6	−13.2	±1.0
400	4.0	−5.0	−10.6	−13.2	±1.0
500	4.1	−5.0	−10.6	−13.2	±1.0
630	4.2	−4.9	−10.5	−13.4	±1.0
800	4.2	−4.8	−10.5	−13.4	±1.0
1000	4.1	−4.8	−10.4	−12.9	±1.0
1250	3.9	−4.8	−10.2	−12.7	±1.0
1600	3.8	−4.8	−10.0	−12.7	±1.0
2000	3.6	−4.7	−10.0	−12.7	±1.0
2500	3.5	−4.6	−9.4	−12.3	±1.0
3150	3.6	−4.6	−9.4	−12.0	±1.0
4000	3.7	−4.6	−9.7	−12.3	±1.5
5000	3.7	−4.5	−9.7	−12.6	±1.5
6300	3.8	−4.5	−9.7	−12.6	±1.5
8000	3.8	−4.9	−10.0	−12.7	±1.5

TABLE 4b/P.51

Normalized free-field response at points on axis in the near field

Frequency	Measurement point						
	5 [a)]	6	7	8	9	10	Tolerance
(Hz)	(dB)	(dB)	(dB)	(dB)	(dB)	(dB)	(dB)
100	5.2	−1.7	−1.4	−4.0	−1.6	−4.2	±1.5
125	5.2	−1.7	−1.3	−3.8	−1.5	−4.2	±1.5
160	5.2	−1.7	−1.2	−3.8	−1.5	−4.2	±1.5
200	5.2	−1.7	−1.2	−3.8	−1.5	−4.2	±1.5
250	5.2	−1.8	−1.3	−3.8	−1.4	−4.2	±1.5
315	5.1	−1.8	−1.3	−3.8	−1.3	−4.2	±1.0
400	5.1	−1.8	−1.3	−3.8	−1.3	−4.0	±1.0
500	5.0	−1.6	−1.3	−3.8	−1.3	−3.9	±1.0
630	5.0	−1.6	−1.3	−3.8	−1.3	−3.9	±1.0
800	5.0	−1.6	−1.3	−3.8	−1.3	−4.0	±1.0
1000	4.8	−1.7	−1.3	−3.9	−1.3	−4.1	±1.0
1250	4.8	−1.8	−1.4	−4.0	−1.3	−4.3	±1.0
1600	4.7	−1.8	−1.4	−3.8	−1.3	−4.0	±1.0
2000	4.7	−1.8	−1.2	−3.7	−1.3	−3.6	±1.0
2500	4.7	−1.9	−1.0	−3.6	−1.1	−3.5	±1.0
3150	4.7	−2.1	−1.1	−3.5	−1.2	−3.4	±1.0
4000	4.5	−2.9	−1.5	−4.1	−1.3	−3.0	±1.5
5000	3.8	−3.6	−1.5	−4.8	−1.3	−3.7	±1.5
6300	3.2	−4.8	−1.8	−5.2	−1.7	−3.7	±1.5
8000	2.5	−5.2	−2.0	−6.1	−2.2	−4.2	±1.5

a) The measurements on the human mouth at point 5 are quite scattered, so the response at this point is only indicatively provided and no tolerances are specified.

TABLE 4c/P.51

Normalized free field response in the far field

Measurement point	Frequency range 100 Hz-8 kHz	
	Response (dB)	Tolerance (dB)
11	−24.0	± 3.0
12	−24.0	± 3.0
13	−24.0	± 3.0
14	−24.0	± 3.0
15	−24.0	± 3.0
16	−24.0	± 3.0
17	−24.0	± 3.0

TABLE 5/P.51

Recommended microphone types for free-field measurements

Measurement point	Microphone size (max.)	Microphone equalization
1, 2, 5, 6, 7, 8, 9, 10	1/4″	Pressure
3, 4	1/2″	Pressure
11, 12, 13, 14, 15, 16, 17	1″	Free-field
MRP	1/4″	Pressure

2.3.2 *Normalized obstacle diffraction*

The normalized obstacle diffraction of the artificial mouth is defined at three points on the references axis, as specified in Table 6/P.51.

Note – If a compressor microphone is used with the mouth, it (or an equivalent dummy) shall be left in place while checking the normalized obstacle diffraction.

2.3.3 *Maximum deliverable sound pressure level*

The artificial mouth shall be able to deliver steadily the acoustic artificial voice at sound pressure levels up to at least +6 dBPa at the MRP.

2.3.4 *Harmonic distortion*

When delivering sine tones, with amplitudes up to +6 dBPa at the MRP, the harmonic distortion of the acoustic signal shall comply with the limits specified in Table 7/P.51.

TABLE 6/P.51

Normalized obstacle diffraction

Frequency (Hz)	Measurement point 18 (dB)	19 (dB)	20 (dB)	Tolerance (dB)
100	32.2	27.0	21.7	±2.0
125	32.0	27.0	21.4	±2.0
160	32.0	27.3	21.4	±2.0
200	31.2	26.5	20.6	±2.0
250	31.2	26.5	20.5	±2.0
315	31.9	27.0	21.0	±1.5
400	31.8	27.0	20.9	±1.5
500	31.3	26.4	20.4	±1.5
630	31.0	26.0	20.0	±1.5
800	30.1	25.1	19.4	±1.5
1000	29.3	24.4	18.8	±1.5
1250	29.0	24.3	18.8	±1.5
1600	28.9	24.5	19.6	±1.5
2000	28.6	25.2	20.5	±1.5
2500	29.0	26.3	23.2	±1.5
3150	29.0	26.5	21.8	±1.5
4000	29.6	27.3	22.8	±2.0
5000	31.2	26.9	22.4	±2.0
6300	31.7	26.0	22.5	±2.0
8000	30.0	23.0	18.0	±2.0

TABLE 7/P.51

Maximum harmonic distortion of the artificial mouth

	Harmonic distorsion 2nd harmonic	3rd harmonic
100 Hz-125 Hz	< 10%	< 10%
125 Hz-200 Hz	< 4%	< 4%
200 Hz-8 Hz	< 1%	< 1%

2.3.5 *Linearity*

A positive or negative variation of 6 dB of the feeding electrical signal shall produce corresponding variation of 6 dB ± 0.5 dB at the MRP for outputs in the range −14 dBPa to +6 dBPa. This requirement shall be met both for complex excitations, such as the artificial voice, and for sine tones in the range 100 Hz to 8 kHz.

2.4 *Miscellaneous*

2.4.1 *Delivery conditions*

The artificial mouth shall be delivered by the maufacturer with the mechanical fixtures required to place the ½″ calibration microphone at the MRP, as specified in Recommendation P.64. Suitable markings shall be engraved on the device housing for identifying the vertical plane position.

Each artificial mouth shall be delivered with a calibration chart specifying the free-field radiation and obstacle diffraction characteristics as defined in this Recommendation

2.4.2 *Stability*

The device shall be stable and reproducible.

2.4.3 *Stray magnetic field*

Neither the d.c. nor the a.c. magnetic stray fields generated by the artificial mouth shall neither influence the signal transduced by microphones under test.

It is recommended that the a.c. stray field produced at the MRP shall lie below the curve formed by the following coordinates:

Frequency (Hz)	*Magnetic output* (dB A/m/Pa)
200	−10
1 000	−40
10 000	−40

It is also recommended that the d.c. stray field at the MRP be lower than 400 A/m.

Note – The recommended d.c. stray field limit of 400 A/m applies specifically to mouths intended for measuring electromagnetic microphones. For measuring other kinds of microphones, a higher limit of 1200 A/m is acceptable.

2.4.4 *Choice of model*

The results of measurements made on the BK 4219 source (no longer produced) and on the newer BK 4227, with its mouthpiece replaced by the UA 0899 conical adaptor, show a satisfactory agreement between the two models and compliance with the present Recommendation. The models actually used in tests shall always be stated, together with the results of measurements.

Note – It should be noted that the BK 4227 artificial mouth generates a d.c. stray magnetic field at the MRP which exceeds 400 A/m. It is then not suitable for measuring electromagnetic microphones.

References

[1] International Electrotechnical Commission Recommendation, *An artificial ear of the wideband type for the calibration of earphones used in audiometry*, IEC Publication 318, Geneva, 1970.

[2] International Electrotechnical Commission Recommendation, *Occluded ear simulator for the measurement of earphones coupled to the ear by ear insert*, IEC Publication 711, Geneva, 1981.

VOLUME METERS

The CCITT considers that, in order to ensure continuity with previous practice, it is not desirable to modify the specification of the volume meter of the ARAEN employed at the CCITT Laboratory.

Table 1/P.52 gives the principal characteristics of various measuring devices used for monitoring the volume or peak values during telephone conversations or sound-programme transmissions.

The measurement of active speech level is defined in Recommendation P.56. Comparison of results using the active speech level meter and some meters described in this Recommendation can be found in Supplement No. 18.

Note – Descriptions of the following devices are contained in the Supplements to *White Book*, Volume V:

- ARAEN volume meter or speech voltmeter: Supplement No. 10 [1].
- Volume meter standardized in the United States of America, termed the "VU meter": Supplement No. 11 [2].
- Peak indicator used by the British Broadcasting Corporation: Supplement No. 12 [3].
- Maximum amplitude indicator Types U 21 and U 71 used in the Federal Republic of Germany: Supplement No. 13 [4].

The volume indicator, SFERT, which formerly was used in the CCITT Laboratory is described in [5].

Comparative tests with different types of volume meters

A note which appears in [6] gives some information on the results of preliminary tests conducted at the SFERT Laboratory to compare the volume indicator with different impulse indicators.

The results of comparative tests made in 1952 by the United Kingdom Post Office appear in Supplement No. 14 [7]. Further results can be found in Supplement No. 18 of the present volume.

TABLE 1/P.52

Principal characteristics of the various instruments used for monitoring the volume or peaks during telephone conversations or sound-programme transmissions

Type of instrument	Rectifier characteristic (see Note 3)	Time to reach 99% of final reading (milliseconds)	Integration time (milliseconds) (see Note 4)	Time to return to zero (value and definition)
(1) "Speech voltmeter" United Kingdom Post Office Type 3 (S.V.3) identical to the speech power meter of the l'ARAEN	2	230	100 (approx.)	equal to the integration time
(2) VU meter (United States of America) (see Note 1)	1.0 to 1.4	300	165 (approx.)	equal to the integration time
(3) Speech power meter of the "SFERT volume indicator"	2	around 400 to 650	200	equal to the integration time
(4) Peak indicator for sound-programme transmissions used by the British Broadcasting Corporation (BBC Peak Programme Meter) (see Note 2)	1		10 (see Note 5)	3 seconds for the pointer to fall to 26 dB
(5) Maximum amplitude indicator used by the Federal German Republic (type U 21)	1	around 80	5 (approx.)	1 or 2 seconds from 100% to 10% of the reading in the steady state
(6) OIRT – Programme level meter: type A sound meter type B sound meter		for both types: less than 300 ms for meters with pointer indication and less than 150 ms for meters with light indication	 10 ± 5 60 ± 10	for both types: 1.5 to 2 seconds from the 0 dB point which is at 30% of the length of the operational section of the scale

Note 1 – In France a meter similar to the one defined in line (2) of the table has been standardized.

Note 2 – In the Netherlands a meter (type NRU-ON301) similar to the one defined in line (4) of the table has been standardized.

Note 3 – The number given in the column is the index n in the formula $[V_{(output)} = V_{(input)}{}^{n}]$ applicable for each half-cycle.

Note 4 – The "integration time" was defined by the CCIF as the "minimum period during which a sinusoidal voltage should be applied to the instrument for the pointer to reach to within 0.2 neper or nearly 2 dB of the deflection which would be obtained if the voltage were applied indefinitely". A logarithmic ratio of 2 dB corresponds to a percentage of 79.5% and a ratio of 0.2 neper to a percentage of 82%.

Note 5 – The figure of 4 milliseconds that appeared in previous editions was actually the time taken to reach 80% of the final reading with a d.c. step applied to the rectifying/integrating circuit. In a new and somewhat different design of this programme meter using transistors, the performance on programme remains substantially the same as that of earlier versions and so does the response to an arbitrary, quasi-d.c. test signal, but the integration time, as here defined, is about 20% greater at the higher meter readings.

Note 6 – In Italy a sound-programme meter with the following characteristics is in use:

Rectifier characteristic: 1 (see Note 3).
Time to reach 99% of final reading: approx. 20 ms.
Integration time: approx. 1.5 ms.
Time to return to zero: approx. 1.5 s from 100% to 10% of the reading in the steady state.

References

[1] *ARAEN volume meter or speech voltmeter*, White Book, Vol. V, Supplement No. 10, ITU, Geneva, 1969.

[2] *Volume meter standardized in the United States of America, termed VU meter*, White Book, Vol. V, Supplement No. 11, ITU, Geneva, 1969.

[3] *Modulation meter used by the British Broadcasting Corporation*, White Book, Vol. V, Supplement No. 12, ITU, Geneva, 1969.

[4] *Maximum amplitude indicators, types U 21 and U 71 used in the Federal Republic of Germany*, White Book, Vol. V, Supplement No. 13, ITU, Geneva, 1969.

[5] *SFERT volume indicator*, Red Book, Vol V, Annex 18, Part 2, ITU, Geneva, 1962.

[6] CCIF *White Book*, Vol. IV, pp. 270-293, ITU, Bern, 1934.

[7] *Comparison of the readings given on conversational speech by different types of volume meter*, White Book, Vol. V, Supplement No. 14, ITU, Geneva, 1969.

Recommendation P.53

PSOPHOMETERS (APPARATUS FOR THE OBJECTIVE MEASUREMENT OF CIRCUIT NOISE)

Refer to Recommendation O.41, CCITT Blue Book, Volume IV, Fascicle IV.4

Recommendation P.54

SOUND LEVEL METERS (APPARATUS FOR THE OBJECTIVE MEASUREMENT OF ROOM NOISE)

(amended at Mar del Plata, 1968 and Geneva, 1972)

The CCITT recommends the adoption of the sound level meter specified in [1] in conjunction, for most uses, with the octave, half, and third octave filters in accordance with [2].

References

[1] International Electrotechnical Commission Standard, *Sound level meters*, IEC Publication 651 (179), Geneva, 1979.

[2] International Electrotechnical Recommendation, *Octave, half-octave and third-octave band filters intended for the analysis of sounds and vibrations*, IEC Publication 225, Geneva, 1966.

Recommendation P.55

APPARATUS FOR THE MEASUREMENT OF IMPULSIVE NOISE

(Mar del Plata, 1968)

Experiments have shown that clicks or other impulsive noises which occur in telephone calls come from a number of sources, such as faulty construction of the switching equipment, defective earthing at exchanges and electromagnetic couplings in exchanges or on the line.

There is no practical way of assessing the disturbing effect of isolated pulses on telephone calls. A rapid succession of clicks is annoying chiefly at the start of a call. It is probable that these series of clicks affect data transmission more than they do the telephone call and that connections capable of transmitting data, according to the noise standards now under study, will also be satisfactory for speech transmission.

In view of these considerations, the CCITT recommends that Administrations use the impulsive noise counter defined in Recommendation O.71 [1] for measuring the occurrence of series of pulses on circuits for both speech and data transmission.

Note – At the national level, Administrations might continue to study whether the use of this impulsive noise counter is sufficient to ensure that the conditions necessary to ensure good quality in telephone connections are met. In those studies, Administrations may use whatever measuring apparatus they consider most suitable – for example a psophometer with an increased overload factor – but the CCITT does not envisage recommending the use of such an instrument.

Reference

[1] CCITT Recommendation *Specification for an impulsive noise measuring instrument for telephone-type circuits*, Vol. IV, Rec. O.71.

Recommendation P.56

OBJECTIVE MEASUREMENT OF ACTIVE SPEECH LEVEL

(Melbourne, 1988)

1 Introduction

The CCITT considers it important that there should be a standardized method of objectively measuring speech level, so that measurements made by different Administrations may be directly comparable. Requirements of such a meter are that it should measure active speech level and should be independent of operator interpretation.

In this Recommendation, a meter is a complete unit that includes the input circuitry, filter (if necessary), processor and display. The processor includes the algorithm of the detection method.

In its present form, this meter can safely be used for laboratory experiments or can be used with care on operational circuits. Further study is continuing on:

a) how the meter can be used on 2-wire and 4-wire circuits to determine who is talking and whether it is an echo, and

b) how such an instrument can discriminate between speech and signalling, for example.

The method described herein maintains maximum comparability and continuity with past work, provided suitable monitoring is used, e.g. an operator performing the monitoring function. In particular, the new method yields data and conclusions compatible with those that have established the conventional value (22 microwatts) of speech power at the input to the 4-wire point of the international circuit according to Recommendation G.223. A method using operator monitoring can be found in Annex A.

This Recommendation describes a method that can be easily implemented using current technology. It also acts as a reference against which other methods can be compared. The purpose of this Recommendation is not to exclude any other method but to ensure that results from different methods give the same result.

Active speech level shall be measured and reported in decibels relative to a stated reference according to the methods described below, namely,

– *Method A* – measuring a quantity called speech volume, used for the purpose of real-time control of speech level (see § 4);

– *Method B* – measuring a quantity called active speech level, used for other purposes (see § 5).

Comparison of readings given by meters of methods A and B can be found in Supplement No. 18.

Note – This meter cannot be used to determine peak levels but sufficient information exists [1] giving the instantaneous peak/r.m.s. ratio, provided the signal has not been restricted or modified in any way, e.g. peak clipping.

2 Terminology

The recommended terminology is as follows:

speech volume	until now used interchangeably with *speech level*, should in future be used exclusively to denote a value obtained by method A;
active speech level	should be used exclusively to denote a value obtained by method B;
speech level	should be used as a general term to denote a value obtained by any method yielding a value expressed in decibels relative to a stated reference.

The definitions of these terms [2], and other related terms such as those for the meters themselves [3], should be adjusted accordingly.

3 General

3.1 *Electrical, acoustic and other levels*

This Recommendation deals primarily with electrical measurements yielding results expressed in terms of electrical units, generally decibels relative to an appropriate reference value such as one volt. However, if the calibration and linearity of the transmission system in which the measurement takes place are assured, it is possible to refer the result backwards or forwards from the measurement point to any other point in the system, where the signal may exist in some non-electrical form (e.g., acoustical). Power is proportional to squared voltage in the electrical domain, squared sound pressure in the acoustical domain, or the digital equivalent of either of these in the numerical domain, and the reference value must be of the appropriate kind (1 volt, 1 pascal, reference acoustic pressure equal to 20 micropascals, or any other stated unit, as the case may be).

3.2 *Universal requirements*

For speech-level measurements of all types, the information reported should include: the designation of the measuring system, the method used (A, B, or B-equivalent as explained in § 4, or other specified method), the quantity observed, the units, and other relevant information such as the margin value (explained below) where applicable.

All the relevant conditions of measurement should also be stated, such as bandwidth, position of the measuring instrument in the communication circuit, and presence or absence of a terminating impedance. Apart from the stated band limitation intended to exclude spurious signals, no frequency weighting should be introduced in the measurement path (as distinct from the transmission path).

3.3 *Averaging*

Where an average of several readings is reported, the method of averaging should be stated. The *mean level* (mean speech volume or mean active speech level), formed by taking the mean of a number of decibel values, should be distinguished from the *mean power*, formed by converting a number of decibel values to units of power, taking the mean of these, and then optionally restoring the result to decibels.

Any correction that has been applied should be mentioned, together with the facts or assumptions on which any such correction is based. For example, in loading calculations, when the active levels or durations of the individually measured portions of speech differ widely, $0.115\ \sigma^2$ is commonly added to the median or mean level in order to estimate the mean power, on the grounds that the distribution of mean active speech levels (dB values) is approximately Gaussian.

4 Method A: immediate indication of speech volume for real-time applications

Measurement of speech volume for rapid real-time control or adjustment of level by a human observer should be accomplished in the traditional manner by means of one of the devices listed in Recommendation P.52.

The choice of meter and the method of interpreting the pointer deflexions should be appropriate to the application, as in Table 1/P.56.

Values obtained by method A should be reported as *speech volume*; the meter employed, the quantity observed, and the units in which the result is expressed, should be stated.

TABLE 1/P.56

Application	Meter	Quantity observed
Control of vocal level in live-speech loudness balances	ARAEN volume meter (SV3)	Level exceeded in 3 s
Avoidance of peak limiting	Peak programme meter	Highest reading
Maintenance of optimum level in making magnetic tape recording	VU meter	Average of peaks (excluding most extreme)

5 Method B: active speech level for other applications than those mentioned in method A

5.1 *Principle of measurement*

Active speech level is measured by integrating a quantity proportional to instantaneous power over the aggregate of time during which the speech in question is present (called the active time), and then expressing the quotient, proportional to total energy divided by active time, in decibels relative to the appropriate reference.

The mean power of a speech signal when known to be present can be estimated with high precision from samples taken at a rate far below the Nyquist rate. However, the all-important question is what criterion should be used to determine when speech is present.

Ideally, the criterion should indicate the presence of speech for the same proportion of time as it appears to be present to a human listener, excluding noise that is not part of the speech (such as impulses, echoes, and steady noise during periods of silence), but including those brief periods of low or zero power that are not perceived as interruptions in the flow of speech [4]. It is not essential that the detector should operate exactly in synchronism with the beginnings and ends of utterances as perceived: there may be a delay in both operating and releasing, provided that the total active time is measured correctly. For this reason, complex real-time voice-activity detectors depending on sampling at the Nyquist rate, such as those that have been successfully used in digital speech interpolation, are not necessarily the most suitable for this application. Their function is to indicate when a channel is available for transmission of information: this state does not always coincide with the absence of speech; on the one hand, it may occur during short intervals that ought to be considered part of the speech, and on the other hand, it may be delayed long after the end of an utterance (for reasons of convenience in the allocation of channels, for example).

This Recommendation describes the detection method that meets the requirements. The method involves applying a signal-dependent threshold which cannot be specified in advance, so that accurate results cannot be guaranteed while the measurement is actually in progress; despite that, by accumulating sufficient information during the process, it is possible to apply the correct threshold retrospectively, and hence to output a correct result almost as soon as the measurement finishes. Continuous adaptation of the threshold level in real time appears to yield similar results in simple cases, but further study is needed to find out how far this conclusion can be generalized.

5.2 *Details of realization*

The algorithm for method B is as follows.

Let the speech signal be sampled at a rate not less than f samples per second, and quantized uniformly into a range of at least 2^{12} quantizing intervals (i.e. using 12 bits per sample including the sign).

Note – This requirement ensures that the dynamic range for instantaneous voltage is at least 66 dB, but two factors combine to make the range of measurable active speech levels about 30 dB less than this:

1) Allowance must be made for the ratio of peak power to mean power in speech, namely about 18 dB where the probability of exceeding that value is 0.001.

2) Envelope values down to at least 16 dB below the mean active level must be calculated: these values may be fractional, but will not be accurate enough if computed from a quantizing interval much exceeding twice the sample value; that is to say, it should not be expected that an active speech level less than about 10 dB above the quantizing interval would be measurable.

Let the successive sample values be denoted by x_i where $i = 1, 2, 3, \ldots$ Let the time interval between consecutive samples be $t = 1/f$ seconds.

Other constants required are:

v	(volts/unit) scale factor of the analogue-digital converter
T	time constant of smoothing in seconds
$g = \exp(-t/T)$	coefficient of smoothing
H	hangover time in seconds
$I = H/t$	rounded up to next integer
M	margin in dB, difference between threshold and active speech level.

Let the input samples be subjected to two distinct processes, 1 and 2.

Process 1

Accumulate the number of samples n, the sum s, and the sum of squares, sq:

$$n_i = n_{i-1} + 1$$

$$s_i = s_{i-1} + x_i$$

$$sq_i = sq_{i-1} + x_i^2$$

where s_0, sq_0 and n_0 (initial values) are zero.

Process 2

Perform two-stage exponential averaging on the rectified signal values:

$$p_i = g \cdot p_{i-1} + (1-g) \cdot |x_i|$$

$$q_i = g \cdot q_{i-1} + (1-g) \cdot p_i$$

where p_0 and q_0 (initial values) are zero.

The sequence q_i is called the envelope, p_i denotes intermediate quantities.

Let a series of fixed threshold voltages c_j be applied to the envelope. These should be spaced in geometric progression, at intervals of not more than 2:1 (6.02 dB), from a value equal to about half the maximum code down to a value equal to one quantizing interval or lower. Let a corresponding series of activity counts a_j, and a corresponding series of hangover counts, h_j, be maintained:

for each value of j in turn,

if $q_i > c_j$ or $q_i = c_j$, then add 1 to a_j and set h_j to 0;

if $q_i < c_j$ and $h_j < I$, then add 1 to a_j and add 1 to h_j;

if $q_i < c_j$ and $h_j = I$, *then* do nothing.

In the first case, the envelope is at or above the jth threshold, so that the speech is active as judged by that threshold level. In the second case, the envelope is below the threshold, but the speech is still considered active because the corresponding hangover has not yet expired. In the third case, the speech is inactive as judged by the threshold level in question.

Initially, all the a_j values are set equal to zero, and the h_j values set equal to I.

It should be noted that the suffix i in all the above cases is needed only to distinguish current values from previous values of accumulated quantities; for example, there is no need to hold more than one value of sq, but this value is continually updated. At the end of the measurement, therefore, the suffixes can be omitted from s, sq, n, p, and q.

Let all these processes continue until the end of the measurement is signalled. Then evaluate the following quantities:

$$\text{Total time} = n \cdot t$$

$$\text{Long-term power} = sq \cdot v^2/n.$$

Note – If it is suspected that there may be a significant d.c. offset, this may be estimated as $s \cdot v/n$, and used to evaluate a more accurate value of long-term power (a.c.) as $v^2\,[sq/n-(s/n)^2]$. However, in this case, the effect of the offset on the envelope must also be taken into account and appropriate corrections made.

For each value of j, the active-power estimate is equal to $sq \cdot v^2/a_j$.

At this stage, the powers are in volts squared per unit time. Now express the long-term power and the active-power estimates in decibels relative to the chosen reference voltage r:

Long-term level, $L = 10 \log (sq \cdot v^2/n) - 20 \log r$

Active-level estimate, $A_j = 10 \log (sq \cdot v^2/a_j) - 20 \log r$

Threshold, $C_j = 20 \log (c_j \cdot v) - 20 \log r$

For each value of j, compare the difference $A_j - C_j$ with the margin M, and determine (if necessary, by interpolation on a decibel scale between two consecutive values of A_j and of C_j) the true active level A and corresponding threshold C for which $A - C = M$. If one of the pairs of values A_j and C_j fulfils this condition exactly, then the true activity factor is a_j/n, but in all cases it can be evaluated from the expression $10^{(L-A)/10}$.

For simplicity, the algorithm has been defined in terms of a digital process, but any equivalent process (one implemented on a programmable analogue computer, for example) should also be considered as fulfilling the definition.

5.3 *Values of the parameters*

The values of the parameters given in Table 2/P.56 should be used. They have been found suitable for the purpose and have stood the test of many years of application by various organizations [4].

TABLE 2/P.56

Parameter	Value	Tolerance
f	694 samples/second	not less than 600
T	0.03 seconds	± 5%
H	0.2 seconds	± 5%
M	15.9 dB	± 0.5

Note – The value M = 15 dB might appear to be implied in [4], but the threshold level there described equals the *mean absolute voltage* of a sine wave whose *mean power* is 15 dB below the reference. The difference of 0.9 dB is 20 log (voltage/mean absolute voltage) for a sine wave.

The result of a measurement made by means of the above algorithm with parameter values conforming to the above restrictions should be reported as *active speech level*, and the system should be described as *using method B* of this Recommendation.

Note – Where noise levels are very high, as they are for example in certain vehicles or in certain radio systems, it is often desirable to set the threshold higher (i.e. use a smaller margin) in order to exclude the noise. This may be done provided the margin is also reported. The result of a such a measurement should be reported as *active speech level with margin M*, and the measurement system described as *using method B with margin M*.

The activity factor should preferably be reported as a percentage, with a specification of the margin value if this is outside the standard range.

6 Approximate equivalents of method B

Other methods under development use a broadly similar principle of measurement but depart in detail from the algorithm given above.

It is not the intention to exclude any such method, provided it is convincingly shown by experimental evidence to yield results consistent with those obtained by method B in a sufficiently wide range of conditions. For this reason, a class of methods called *B-equivalent methods* is recognized.

A B-equivalent method of speech-level measurement is defined as any method that satisfies the following test in all respects.

Measurements shall be carried out simultaneously by the method in question and by method B on two or more samples of speech in every combination of the following variables:

Voices	one male and one female voice
Speech material	a list of independent sentences, a passage of continuous speech, and one channel of a conversation, each lasting at least 20 s (active time)
Bandwidth	300 to 3400 Hz and 100 to 8000 Hz
Added noise	flat within the measurement band at levels $(M + 5)$ dB and $(M + 25)$ dB below the active speech level, where M (the margin) is normally 15.9 dB, but smaller in high-noise applications
Levels	at intervals of 10 dB over the range claimed for the system in question.

From the results, 95% confidence limits for the difference between the level given by the method in question and the active speech level given by method B shall be calculated for each of the above 24 combinations.

If, for every combination, the upper confidence limit of this difference is not higher than +1 dB and the lower confidence limit is not lower than −1 dB, then the method shall be deemed to be a B-equivalent method.

This verification procedure is valid until a suitable speech-like signal has been recommended and found suitable to perform this function (see Questions 12/XII and 13/XII).

Further, a method qualifies as B-equivalent if it gives results that fall within the specified limits when corrected by the addition of a fixed constant, known in advance of the measurement and not dependent on any feature of the speech signal (except possibly the bandwidth if this is known independently).

The results of measurements by such a method should be reported as *B-equivalent active speech level*, and the activity factor as *B-equivalent activity factor*.

Certain measurement systems with fixed thresholds (instead of the retrospectively selected threshold as described in § 5.3), may still give an active speech level according to the definition in cases where the margin turns out to be within the specified limits.

7 Specification

A speech voltmeter normally consists of three parts, namely:

i) input circuitry,

ii) filter, and

iii) processor and display.

Figure 1/P.56 shows a typical layout of such a meter.

Whether all or part of the components that make up i) and ii) are used will depend on where the meter is to be used. However, it is recommended that a meter for general usage should conform to this specification.

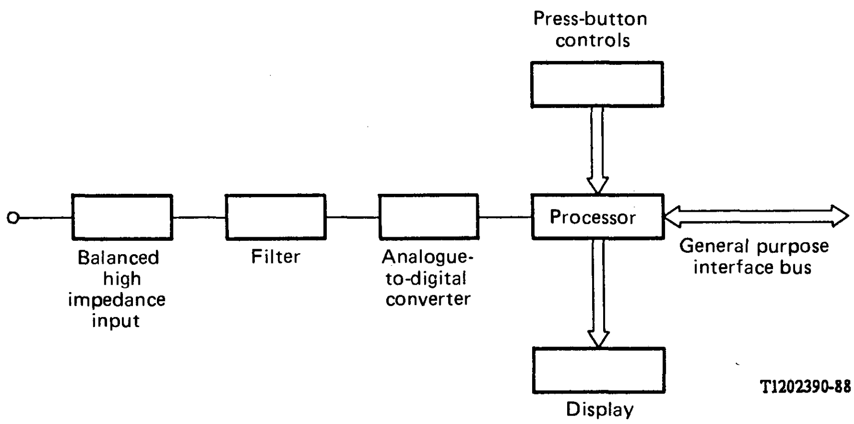


FIGURE 1/P.56

Block diagram of a typical meter

7.1 *Signal input*

7.1.1 *Input impedance*

The meter is normally used as a bridging instrument and, if so, its impedance must be high so as not to influence the results. An impedance of 100 kohm is recommended.

7.1.2 *Circuit protection*

It is recommended that the meter should withstand voltages far in excess of those in the measurement range as accidental usage may occur and the circuit under test may have higher voltages than anticipated. Examples of this are mains 110/240 V or 50 V exchange voltages.

7.1.3 *Connection*

It is recommended that the connection should be independent of polarity. The meter should have the facility of connection in both balanced and unbalanced modes.

7.2 *Filter*

When measuring the speech levels of circuits in the conventional telephony speech bandwidth (300-3400 Hz), it is often practical to use a filter that will reject unwanted hum, tape noise, etc. yet pass the frequencies of greatest interest without affecting the speech level measurement. The set of coordinates in Table 3/P.56 meet these requirements. Figure 2/P.56 gives an example of such a filter.

The following noise requirements should also be met:

Output noise level:

wideband (20-20 000 Hz) < -75 dBm

telephone weighted < -90 dBmp.

TABLE 3/P.56

Frequency (Hz)	(dB)
	Upper limit response relative to 1 kHz
16	−49.75
160	+0.25
7 000	+0.25
70 000	−49.75
	Lower limit response relative to 1 kHz
Under 200	−∞
200	−0.25
5500	−0.25
Over 5500	−∞

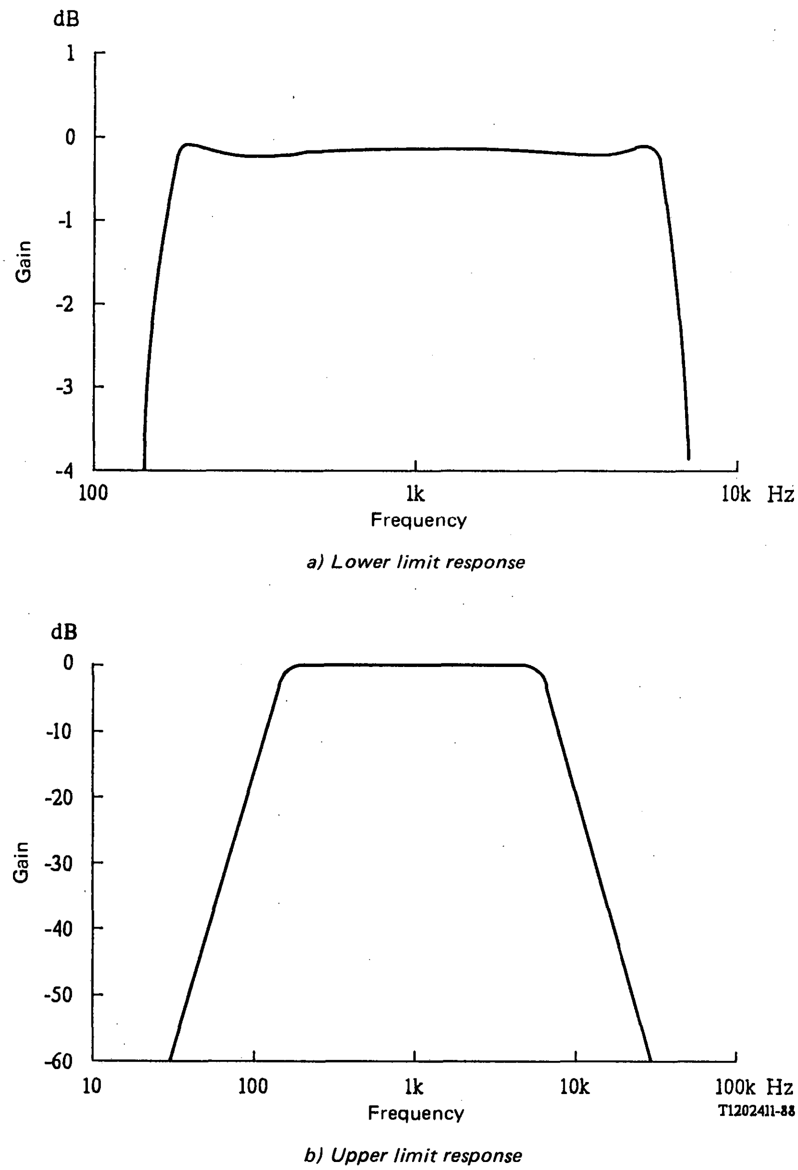


FIGURE 2/P.56

Filter passband response

7.3 *Speech level measurements*

7.3.1 *Working range for speech*

The recommended working range for speech refers to the active level and should be at least 0 to −30 dBV.

Note 1 – The dynamic range of the instrument will depend on the analogue-to-digital converter (ADC). If the ADC is set to a 10 volt maximum input level (i.e. the all 1 code) and 12-bit arithmetic is used, based on the most significant bits from the ADC, then 1 sign bit +11 bits magnitude provides a 66 dB range. The measurable range sill be some 35 dB less when allowance is made for the peak/mean ratio of 18 dB (peaks of speech will only exceed the maximum input level for less than 0.1% of the time [1]) and margin M of 15.9 dB; the largest speech signal is therefore around +2 dBV with a smallest speech signal of −30 dBV. However, the practical working range has been found to be +5 dBV to −35 dBV.

Note 2 – To cater for a wider range of speech levels, an attenuator or low noise amplifier may be inserted in the input circuitry. Care must be exercised to maintain the input requirements of § 7.1.1.

7.3.2 *Linearity*

The linearity of the meter is specified for r.m.s. sine wave measurements since for speech the algorithm is correct by definition, and only the precision or repeatability of measurements need to be considered; this is specified in § 7.3.4.

Assuming that:

a) the measurement is for a minimum period of 5 s,

b) the sine wave is present for the whole of the measurement period, the linearity specified is:

Frequency (Hz)	*Input range* (dBV)	*Accuracy* (dB)
100 to 4000	+16 to −45	± 0.1
4000 to 8000	+13 to −45	± 0.3

Note – The maximum input for the frequency range 4000 to 8000 Hz should ideally be the same as for 100 to 4000 Hz, but practical limitations in commercially available ADCs (due to the limited "slewing rate" of the input circuitry) means that this cannot be obtained. However, as the power in the 8000 Hz band for speech is 30 dB down on the level at 500 Hz it is likely that any error will be extremely small.

7.3.3 *Frequency response*

The frequency response of the meter without filter when measured in the frequency range 100 to 8000 Hz should be flat within the specified tolerances:

Frequency (Hz)	*Input range* (dBV)	*Tolerance* (dB)
100 to 4000	+16 to −45	± 0.2
4000 to 8000	+13 to −45	± 0.4

Note 1 – Tolerances are referred to 1000 Hz.

Note 2 – The note of 7.3.2 applies.

7.3.4 *Repeatability*

When a given speech signal, having its active level within the recommended working range and its duration not less than 5 s active time, is repeatedly measured on the same meter, the active-level readings shall have a standard deviation of less than 0.1 dB.

8 Routine calibration of method-B meter

The following routine calibration procedures, using non-speech-like signals, will ensure that the meter is performing satisfactorily. The calibration can only be made using speech.

A suitable circuit arrangement is shown in Figure 3/P.56. Wherever suitable, measurements should be made with two settings of the attenuator, 0 and 20 dB. All source signals are from a 600 ohm source and the meter is terminated in 600 ohm.

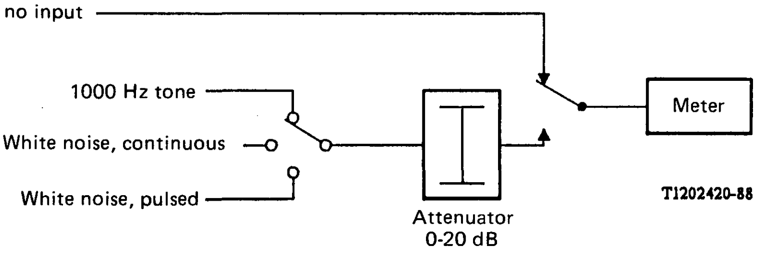


FIGURE 3/P.56

Switching arrangement

8.1 *No input signal*

With no input applied the meter should display the following results:

Activity factor	0 + 0.5%
Active-level	< −60 dBV
Long-term level	< −60 dBV

8.2 *Continuous tone*

With a 1000 Hz sine wave calibrated to be 0 dBV, the meter should display the following results for the two settings of the attenuator when applied for 12 + 0.2 s:

	Attenuator = 0 dB	*Attenuator = 20 dB*
Activity factor	100 to 0.5%	100 to 0.5%
Active-level	0 ± 0.1 dBV	−20 ± 0.1 dBV
Long-term level	0 ± 0.1 dBV	−20 ± 0.1 dBV

8.3 *White noise*

8.3.1 *Without filter*

With the meter having no filter in circuit and the white noise source calibrated to be 0 dBV, the meter should display the following results for the two settings of the attenuator when applied for 12 + 0.2 s:

	Attenuator = 0 dB	*Attenuator = 20 dB*
Activity factor	100 to 0.5%	100 to 0.5%
Active-level	0 ± 0.5 dBV	−20 ± 0.5 dBV
Long-term level	0 ± 0.5 dBV	−20 ± 0.5 dBV

8.3.2 *With filter*

With the meter having the filter in circuit and the white noise source calibrated to be 0 dBV, the meter should display the following results for the two settings of the attenuator when applied for 12 + 0.2 s:

	Attenuator = 0 dB	*Attenuator = 20 dB*
Activity factor	100 to 0.5%	100 to 0.5%
Active-level	-6.9 ± 0.5 dBV	-26.9 ± 0.5 dBV
Long-term level	-6.9 ± 0.5 dBV	-26.9 ± 0.5 dBV

8.3.3 *Pulsed noise*

With the meter having no filter in circuit and the white noise source pulsed at 3 s "ON" and 3 s "OFF" and calibrated to be 0 dBV when "ON", the meter should display the following results for the two settings of the attenuator when applied for 12 + 0.2 s:

	Attenuator = 0 dB	*Attenuator = 20 dB*
Factor activity	55 ± 1.5%	55 ± 1.5%
Active-level	0 ± 1 dBV	-20 ± 1 dBV
Long-term level	-2.7 ± 1 dBV	-22.7 ± 1 dBV

Note – It is possible that § 8 could be revised to calibrate both method B and B-equivalent meters when a speech-like signal has been found suitable to perform this function.

ANNEX A

(to Recommendation P.56)

A method using a speech voltmeter complying with method B in network conditions

A speech voltmeter complying with method B is not suitable in its present form for speech measurements (see, for example, Recommendation G.223) on real connections since the meter is unable to distinquish between speech coming from one or the other end of the connection.

However, if the meter is connected to a 4-wire point in a connection of the type 2-4-2 wire, then measurements may be made using an operator monitoring the beginning and the end of the conversation. The operator can perform this function using earphones (provided the subscriber's permission has been obtained) or by an auxiliary meter (for example conforming to P.52). The circuit arrangement is shown in Figure A-1/P.56.

The operator monitors the conversation, using the auxiliary meter or earphones, and then by means of a start/stop button can measure the beginning and end of the relevant conversation.

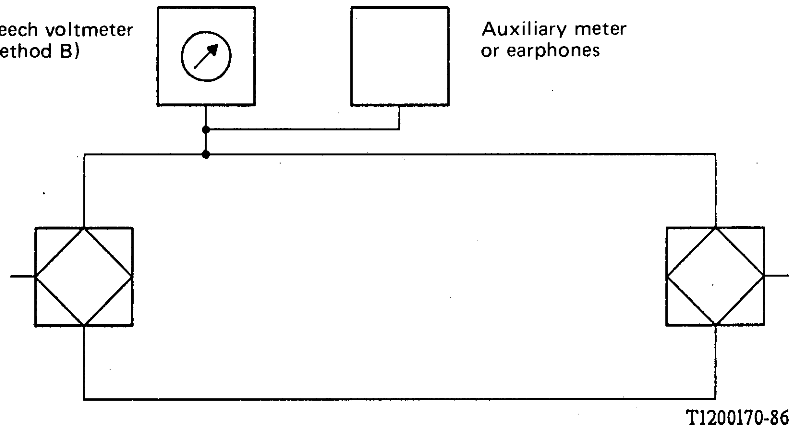


FIGURE A-1/P.56

References

[1] RICHARDS (D. L.): Telecommunication by speech, § 2.1.3.2, pp. 56-69, *Butterworks*, London, 1973.

[2] ITU – *List of Definitions of Essential Telecommunication Terms*, Definition 14.16, Second impression, Geneva, 1961.

[3] ITU – *List of Definitions of Essential Telecommunication Terms*, Definitions 12.34, 12.35, 12.36, Second impression, Geneva, 1961.

[4] BERRY (R. W.): Speech-volume measurements on telephone circuits, *Proc. IEE*, Vol. 118, No. 2, pp. 335-338, February 1971.

Bibliography

BRADY (P. T.): Equivalent Peak Level: a threshold-independent speech level measure, *Journal of the Acoustical Society of America*, Vol. 44, pp. 695-699, 1968.

CARSON (R.): A digital Speech Voltmeter – the SV6, *British Telecommunications Engineering*, Vol. 3, Part 1, pp. 23-30, April 1984.

CCITT – Contribution COM XII-No. 43 *A method for speech-level measurements using IEC-interface bus and calculation* (Norway), Geneva, 1982.

PAGE INTENTIONALLY LEFT BLANK

PAGE LAISSEE EN BLANC INTENTIONNELLEMENT

SECTION 5

OBJECTIVE ELECTRO-ACOUSTICAL MEASUREMENTS

Recommendation P.61

METHODS FOR THE CALIBRATION OF CONDENSER MICROPHONES

(amended at Malaga-Torremolinos, 1984)

Primary and secondary calibrations of condenser microphones can be carried out using the methods described below.

1 Primary calibration by the reciprocity method

The recommended procedure for primary calibration of condenser microphones is the reciprocity calibration technique. A precision method for reciprocity pressure calibration is described in [1]. A simplified method, suitable for calibration over the frequency range of interest for telephonometric measurements, is given in [2]. Although the methods described are specifically for one-inch microphones, similar methods are applicable to half-inch microphones. Methods suitable for half-inch microphones are under study by IEC.

A precision method for free-field reciprocity calibration is given in [3]. Alternatively, the free-field correction curves given in [4] may be applied to the pressure calibration of one-inch condenser microphones to determine their free-field responses. The reciprocity free-field calibration method may in principle be extended to half-inch microphones. Free-field correction curves have not been standardized for half-inch microphones.

2 Secondary calibration by the comparison method

The secondary calibration of a condenser microphone may be achieved by direct comparison with a physically identical microphone having a known calibration. The procedure used is a modification of the "two microphones and auxiliary sound source" method described in [1] to [3]. The output of the calibrated microphone is first determined for a given drive level applied to the auxiliary sound source. The calibrated microphone is then replaced by the microphone to be calibrated, and its output is determined for the same drive level applied to the auxiliary sound source. The difference in level (in dB) between the outputs of the two microphones is then applied to the known calibration of the first microphone to determine the calibration of the second. The procedure is repeated at each frequency of interest.

3 Secondary calibration using pistonphones and other sound level calibrators

Secondary calibrations can also be made using pistonphones and other sound level calibrators which produce a known sound level. Such devices are often used to check the calibration of a microphone at a single frequency [1]. Care should be taken to follow the manufacturer's instructions when using such devices; in particular, it may be necessary to apply corrections for barometric pressure, coupler volume, microphone type, etc. Standardization of these calibrators is currently under study by the IEC.

---

[1] Calibrations with an accuracy of ± 0.3 dB are possible.

References

[1] International Electrotechnical Commission *Precision method for pressure calibration of one-inch standard condenser microphones by the reciprocity technique*, IEC publication 327, Geneva, 1971.

[2] International Electrotechnical Commission *Simplified method for pressure calibration of one-inch condenser microphones by the reciprocity technique*, IEC publication 402, Geneva, 1972.

[3] International Electrotechnical Commission *Precision method for free-field calibration of one-inch standard condenser microphones by the reciprocity technique*, IEC publication 486, Geneva, 1974.

[4] International Electrotechnical Commission *Values for the difference between free-field and pressure sensitivity levels for one-inch standard condenser microphones*, IEC Publication 655, Geneva, 1979.

Recommendation P.62

MEASUREMENTS ON SUBSCRIBERS' TELEPHONE EQUIPMENT

(amended at Malaga-Torremolinos, 1984 and Melbourne, 1988)

1 Measurement of the attenuation distortion of a telephone set

The curve of the variation of the absolute sensitivity of an item of telephone equipment (sending or receiving system) as a function of frequency does not supply complete information on the manner in which this equipment reproduces the human voice or music, although such a curve may often be called the frequency characteristic.

However, the curve of variation of the absolute sensitivity of telephone equipment as a function of frequency gives useful indications from the point of view of the transmission of speech. On the other hand, for the transmission of music, in the absence of a precise criterion of the quality of transmission (corresponding to articulation, or repetition rate, in commercial telephony) such curves should be sufficient to enable the quality of the terminal equipment used (microphone or loudspeakers) to be appreciated.

For tracing sensitivity/frequency characteristics the methods described in Recommendation P.64 and its associated Annex B may be used.

2 Measurement of the nonlinear distortion of a telephone set and of microphone noise

While the nonlinear distortion of telephone receivers is in general negligible, microphones (and particularly carbon microphones of the type generally used in commercial telephone equipment) show considerable nonlinearity: the relationship between the variation of microphone resistance and the acoustic pressure on the diaphragm is not linear. This nonlinearity becomes more important as the variation of resistance in relation to the total resistance of the microphone increases, i.e. when the microphone is more sensitive. Furthermore, there may be two supplementary effects:

1) The microphone is less sensitive to acoustic pressure lower than a certain value (threshold of excitation).

2) As a consequence of the mechanical inertia of the carbon granules (delay in establishing electrical contact between the granules), the various states of agitation of the carbon under the influence of acoustic waves are not the same for all frequencies (for example, slow beats between two sounds are in general enhanced in reproduction by a carbon microphone).

Existing information on the general effect of harmonic distortion on telephone speech quality indicates that the effect of second order distortion is considerably less than that of third order distortion. Absolute detection thresholds obtained in different test are, however, difficult to compare because of differences in definition and measurement of the distortion.

Note 1 – Summaries of information available in this area are given in [1] and [2]. It is clear that measurements with sinusoidal signals can predict the speech transmission performance of nonlinear systems only to a limited extent, particularly if the peak value of the test signal is much smaller than the transmitted speech signal. A complex signal having the same spectral density at the same amplitude density function as real speech, see Recommendation P.50, is therefore expected to be a more useful test signal.

Note 2 – The application of complex test signals or actual speech signals for the measurement of nonlinearity in telephone circuits is studied under Question 13/XII [3].

Certain types of carbon microphones may produce an audible stationary noise, often depending on the size of feeding current. The measurement of this kind of noise and its effect on transmission quality is the same as for other kinds of additive circuit noise.

3 Objective measurement of loudness rating (LR)

Examples of apparatus that objectively measure LRs conforming to Recommendation P.65 are "CERF" of the French Administration [4], "AURAL" of NTT [5], "TIGGER" [6] of British Telecom and "Loudness Rating Meter" [7] of STL. Short descriptions of the apparatus named above can be found in Chapter 5 of the CCITT *Handbook on Telephonometry* [8].

References

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

[2] CCITT – Question 13/XII, Annex, Green Book, Vol. V, ITU, Geneva, 1973.

[3] CCITT – Question 13/XII, Contribution COM XII-No. 1, Study Period 1985-1988, Geneva, 1985.

[4] CCITT – Contribution COM XII-No. 184, *Equipment for the objective measurement of equivalent R25 and of the sidetone – used by the French Administration* (France), Study Period 1981-1984.

[5] CCITT – Contribution COM XII-No. 79, *Objective loudness rating measurement system,* (NTT), Study Period 1981-1984.

[6] WARD (H. F.) and CROSS (R. C.) TIGGER: An Automatic Test System for measuring the Transmission Performance of Telephones, *British Telecommunications Engineering,* Volume 2, July 1983.

[7] CCITT – Question 15/XII, Annex 6, Contribution COM-No. 1, Study Period 1985-1988.

[8] CCITT *Handbook on Telephonometry,* UIT, Geneva, 1987.

Recommendation P.63

METHODS FOR THE EVALUATION OF TRANSMISSION QUALITY ON THE BASIS OF OBJECTIVE MEASUREMENTS

These measuring methods are being studied by the CCITT under Question 7/XII [1]. Annexes A and B to Recommendation P.11 and Supplements No. 2 and 3, at the end of this Fascicle, describe methods used respectively by British Telecom and AT&T. Attention is also drawn to methods for calculating loudness ratings given in Recommendation P.79.

Reference

[1] CCITT – Question 7/XII, Contribution COM XII-No. 1, Study Period 1985-1988, Geneva, 1985.

DETERMINATION OF SENSITIVITY/FREQUENCY CHARACTERISTICS OF LOCAL TELEPHONE SYSTEMS TO PERMIT CALCULATION OF THEIR LOUDNESS RATINGS

(Geneva, 1976; amended at Malaga-Torremolinos, 1984 and Melbourne, 1988)

See Recommendation P.76 for general principles concerning the determination of loudness ratings.

1 Introduction

The sending, receiving or sidetone sensitivity/frequency characteristic of a local telephone system (LTS) is usually measured directly.

Note 1 – The sending, receiving or sidetone sensitivity/frequency characteristic can also be calculated provided the relevant information of the telephone line and feeding bridge is known. Some of the information required for sidetone is outside the scope of the existing Recommendations.

Note 2 – The same principles also apply to the measurement of microphones and earphones.

Since electro-acoustical measurements of the type being considered may be required for different purposes, it is important to distinguish the following:

a) supplying the designer of a transducer with information concerning the success he has achieved in aiming at a given sensitivity/frequency response;

b) checking that the manufactured product meets the specified requirements;

c) supplying sensitivity/frequency characteristics suitable for use in calculating loudness ratings, or estimating other subjectivity-determined quantities.

The present Recommendation is concerned only with c) and, for this purpose, measurements under real conditions must form the basis. Artificial mouths and artificial ears must be used with due regard to obtaining good agreement between these measurements and those from real mouth and ear determinations. Measurements under real conditions are complicated, time-consuming and not reproducible with great precision, especially when carbon microphones are involved.

The present Recommendation describes measurement methods using recommended forms of artificial mouths and artificial ears (see Recommendation P.51).

This Recommendation applies mainly to LTSs with handset telephones. However, the principles also apply to other types of telephones. Specific considerations for headsets are described in Recommendation P.38, and for loudspeaker telephones in Recommendation P.34.

2 Sending sensitivities of the LTS

For the present purposes, the sending sensitivity of a local telephone system is specified in terms of the free-field sound pressure at a reference point in front of the mouth[1], and the electrical output from the local telephone system or the microphone as the case may be. The input sound pressure cannot be measured simultaneously with the electrical output and therefore the measurement must be made in an indirect manner. The sound pressure at the reference point is measured in the absence of the handset and, with the artificial mouth source unchanged, the handset is placed in the defined position in front of the mouth and the output measured. When a human mouth and voice are used, the source cannot be relied upon to maintain its output constant between the measurement of free-field sound pressure and that of the electrical output from the microphone. Artificial mouths suffer from imperfect representation of the source impedance and field distribution that applies to real mouths.

[1] The mouth reference point used in the present Recommendation is defined in Annex A.

In addition to providing the proper source conditions, it is necessary to ensure that the mouthpiece is located for every design of telephone handset at the position that would be used in the real situation. This can be achieved by locating the mouthpiece properly with respect to an ear reference point; this ensures that longer handsets are measured with a greater mouth-to-microphone distance than is the case for shorter handsets. The success of using a given handset measuring position for measurement of sensitivity/frequency characteristics can be judged only by making comparisons, for handsets of different lengths, between real conversation test results using the artificial mouth and real mouths under suitably controlled measuring conditions. For the present Recommendation, the telephone handset shall be located as defined in Annex A of Recommendation P.76.

Special problems are encountered when making measurements with real mouths and real voices, even under controlled talking conditions. Under such circumstances the sound pressure cannot be measured directly at the required mouth reference point and therefore it has to be measured at some other point and referred indirectly to the mouth reference point. Some previous determinations have made use of a measuring microphone 1 metre from the mouth but this requires anechoic surroundings and is affected by obstruction from the handset under test. Other methods have been also tried and none seems satisfactory so far.

When the sound pressure input to a carbon microphone is increased, the corresponding increase in output voltage does not bear a linear relationship to the increase in sound pressure. This nonlinearity is a very complicated function of applied sound pressure, frequency, feeding current, conditioning and granule-chamber orientation. Reproducible results are obtained with an artificial mouth only if proper attention is paid to all these factors.

3 Receiving sensitivities of the LTS

The IEC-318 model artificial ear (see Recommendation P.51) provides means for precise measurements of the receiving sensitivities of the LTS. However, the sound pressures measured with it do not always agree well with those existing at the ear reference point in real ears under the test conditions used when subjective determinations of loudness ratings are being made. This can be attributed partly to the presence of appreciable acoustical leakage (L_E) between the earphone and the real ear (such leakage is not represented in available recommended forms of the artificial ear) and partly to an increase in enclosed volume between the forms of earphone and the forms of real ear. Therefore, to use the results of measurements made according to the present Recommendation, it is necessary to make a correction (see § 7 below).

Clearly, it would be very desirable if the artificial ear could be modified so as to avoid the need for the correction. Some further work has been done on this matter but it is not yet clear whether a single modification to the artificial ear would suffice for all types of telephone earphone. Further evidence is required, preferably from several laboratories so that a much wider variety of types of earphone can be examined.

4 Artificial mouth and voice

The following properties are required:

a) the distribution in sound pressure around the orifice must be a good approximation to that around a human mouth;

b) the acoustical impedance looking into the mouth must simulate that for human mouths, so that the pressure increase caused by the obstruction effect of telephone microphones will be representative;

c) it must be possible to establish definite sound pressures at the mouth reference point as a function of frequency. A convenient feature to embody in a practical artificial mouth is the linearity, over a suitable range of sound pressures, of the ratio of sound pressure at the mouth reference point to the voltage input to the artificial mouth. The ratio must be independent of frequency at least over the range 200 to 4000 Hz but preferably 100 to 8000 Hz.

For the present purposes the mouth reference point (MRP) is defined by the point on the axis of the artificial mouth located 25 mm in front of the equivalent lip position (see Annex A).

Recommendation P.51 defines the requirements for artificial mouths.

Note – However, the send loudness ratings calculated from the sending sensitivities measured when using an artificial mouth do not always agree well with the loudness ratings determined subjectively using real mouths. The subject is still under study in Questions 8/XII and 12/XII.

In principle, the artificial voice defined in Recommendation P.50 should be used as the acoustic test signal. However, sine waves at defined frequencies have been used satisfactorily so far as stable sets are concerned. Some other signals with continuous spectra, for example pink noise and Gaussian noise having the same long-term spectrum as speech, can also be used as the acoustic test signal. Sine waves can also be used for the measurement of some types of carbon microphones if appropriate techniques are used (see Annex B).

5 Artificial ear

The following properties are required:

a) the acoustical impedance presented to telephone earphones must simulate that presented by real ears under practical conditions of use of telephone handsets;

b) the sensitivity of the artificial ear is defined as the pressure sensitivity of the measuring microphone. It should be constant within ± 0.5 dB over the frequency range 100-8000 Hz.

For a human ear, the ear reference point (ERP) is defined in Annex A. The corresponding point when the ear-cap is fitted to an artificial ear will usually differ from the place at which the sound pressure is measured and for this and other reasons certain corrections are necessary when the results are used for calculating loudness ratings (see § 3 above).

6 Definition of sending sensitivity of an LTS

The sending sensitivity of an LTS, depends upon the location of the handset relative to the equivalent lip position of the artificial mouth. For the present purposes the speaking position defined in Annex A to Recommendation P.76 shall be used. Usually, the sending sensitivity is a function of frequency.

The sending sensitivity of a local telephone system at a specified frequency or in a narrow frequency band is expressed as follows:

$$S_{mJ} = 20 \log_{10} \frac{V_J}{p_m} \text{ dB rel 1 V/Pa}$$

where V_J is the voltage across a 600 ohms termination and p_m is the sound pressure at the mouth reference point. Note that p_m must be measured in the absence of the "unknown" handset of the test item.

6.1 *Measurement of telephone sets containing carbon microphones*

It is intended that the Recommendation should apply for measuring systems containing carbon microphones as well as those having noncarbon microphones. When measuring LTSs that contain linear items, it does not matter at which sound pressure the measurements are made as long as it is known and does not cause overloading. However, when carbon microphones are present, different sensitivities will be obtained depending upon the sound pressure and characteristics of the acoustic signal used. For calculation of sending loudness rating, these must be reduced to single values at each frequency and the method of reduction must take account of the characteristics of human speech. At present, there is no single method that can be recommended for universal use. The problem is being studied under Question 8/XII [1]. Until a suitable method can be defined, Administrations may take note of the various methods that have been suggested and are undergoing appraisal; they are indicated in Annex B.

7 Definition of receiving sensitivity of an LTS

Usually, the receiving sensitivity is a function of frequency. The receiving sensitivity of a local telephone system at a specified frequency or in a narrow frequency band, as measured directly with an artificial ear complying with Recommendation P.51, is expressed as follows:

$$S_{Je} = 20 \log_{10} \frac{p_e}{\frac{1}{2} E_J} \text{ dB rel 1 Pa/V}$$

where p_e is the sound pressure in the artificial ear and ½ E_J is half the emf in the 600 ohm source.

Note – The receiving sensitivity suitable for use in calculation of loudness is given by:

$$S_{JE} = S_{Je} - L_E$$

where L_E is a correction explained above in § 3 and S_{JE} is the receiving sensitivity determined using a large number of real ears.

Further information on this topic is given in Recommendation P.79.

8 Definitions of talker and listener sidetone sensitivities of an LTS

The talker sidetone sensitivity of an LTS is a function of the sending and receiving sensitivities of the telephone set, but also depends on a number of factors including the local subscriber's line conditions, the effective terminating impedance at the local exchange and the sidetone balance circuit within the telephone set.

The sidetone sensitivity as measured from an artificial mouth to the telephone earphone is expressed as:

$$S_{meST} = 20 \log_{10} \left(\frac{p_e}{p_m} \right) \quad \text{dB}$$

where p_m is defined in § 6 and p_e is the sound pressure developed in the artificial ear with the handset in the loudness rating guard ring position (LRGP).

The listener sidetone sensitivity as measured in a diffuse room noise field is expressed as:

$$S_{RNST} = 20 \log_{10} \left(\frac{p_e}{p_{RN}} \right) \quad \text{dB}$$

where p_e is the sound pressure developed in the artificial ear with the handset held at LRGP in front of an unenergised artificial mouth, for a diffuse room noise sound pressure p_{RN} measured at the MRP, but in the absence of all obstacles (e.g. test head, handset, etc.).

9 Methods for determining S_{mJ}, S_{Je}, S_{meST}, S_{RNST} and Δ_{SM}

When the sending, receiving and sidetone sensitivities of an actual local telephone system are required, the measurements according to the definitions given in §§ 6, 7 and 8 above can be made as illustrated in Figures 1/P.64, 2/P.64, 3/P.64, 4/P.64 and 5/P.64. These methods have been used by CCITT Laboratory and elsewhere successfully.

When using fast Fourier transform (FFT) techniques for measuring the characteristics of non-linear LTS, the measurement principle used, i.e. ratio of r.m.s. variables, or crosspectrum (coherent) method, should be specified.

More detail may be found in Section 3 of the *Handbook of Telephonometry* [2].

Figure 1/P.64 shows the method of setting up the artificial mouth so that the sound pressure p_m at the mouth reference point is known at each test frequency or frequency band. It is recommended to provide equalization in the artificial mouth drive circuit to maintain the free-field sound pressure constant at the MRP to within ± 1 dB over the frequency range 100 to 8000 Hz. In no case should the deviation exceed ± 2 dB over the frequency range 200 to 4000 Hz and +2/−5 dB over the frequency range 100 to 8000 Hz. It is recommended that any deviations from the desired sound pressure level be taken into account when determining the sending or sidetone sensitivity of a local telephone system. This is particularly true if the deviation exceeds ± 1 dB.

For any test signal, p_m of −4.7 dBPa is recommended (see Note 2 to Annex B for information).

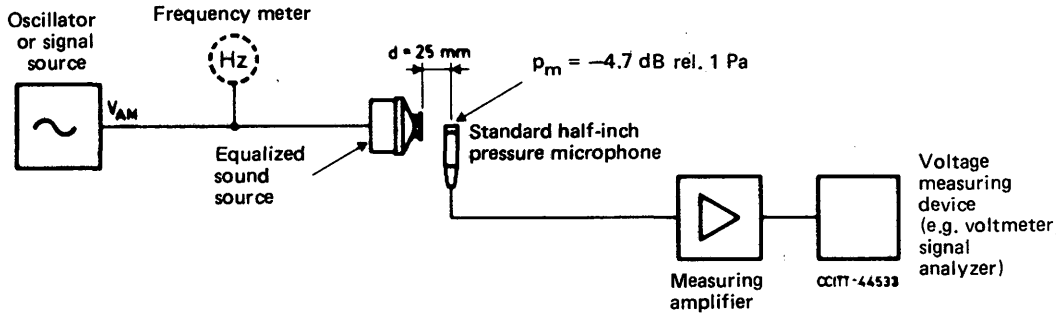


FIGURE 1/P.64

Measurement of acoustic pressure p_m at the mouth reference point 25 mm from the artificial lip plane of the sound source

Figure 2/P.64 shows the measurement of output V_J from the local telephone system when the handset is placed at the appropriate position in front of the artificial mouth and the artificial mouth is energized in the same manner as when the sound pressure p_m was set up in the absence of the handset under test (see Figure 1/P.64).

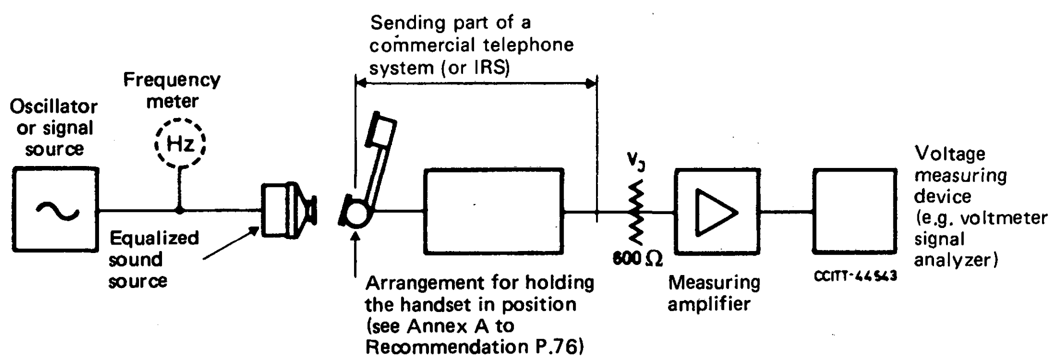


FIGURE 2/P.64

Voltage V_J, measured across the terminals of a 600 ohms pure resistance connected to the output of a sending system under test

Figure 3/P.64 shows the measurement of the sound pressure p_e in the artificial ear when the local telephone system is connected to a 600-ohm source of internal emf E_J. Note that the definition of S_{Je} is in terms of 1/2 E_J and not the potential difference across the input terminals of the local telephone system; this potential difference will, of course, differ from 1/2 E_J, if the input impedance of the local telephone system is not 600 ohms. Care must be taken to ensure that there is no coupling loss (acoustic leakage) between the ear-piece of the receiving system under test and the artificial ear. Usually $E_J = -12$ dBV is recommended.

Note – Some receiving systems incorporate electronic circuits to provide special features, for example, compression to limit the level of the received sound signal. Particular care must be exercised during the measurement of such systems to ensure that the resulting sensitivity is correct and relevant. In some cases it may be necessary to determine the receiving sensitivity over a range of input levels.

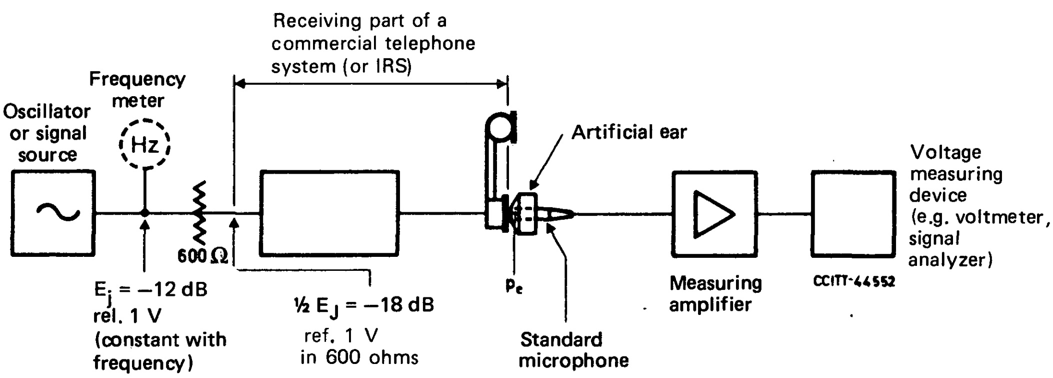


FIGURE 3/P.64

Measurement of acoustic pressure p_e in the artificial ear developed by the receiving system under test

Figure 4/P.64 shows the measurement of sidetone sensitivity. The resulting value of S_{meST} is highly dependent on the impedance connected to the telephone set terminals and therefore, under short line conditions, on the exchange termination. As this impedance often deviates considerably from 600 ohms, particularly when there is a complete connection present, 600 ohms is given only as an example.

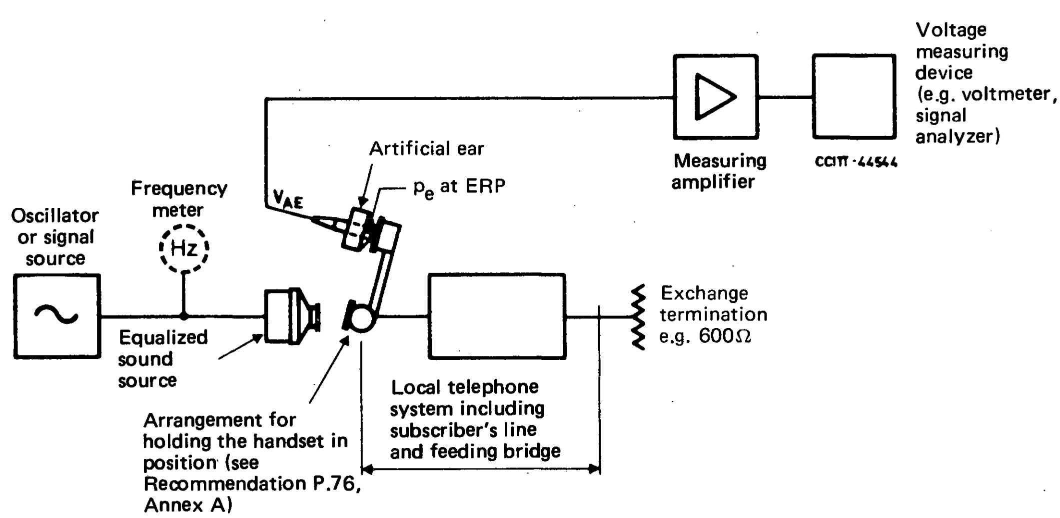
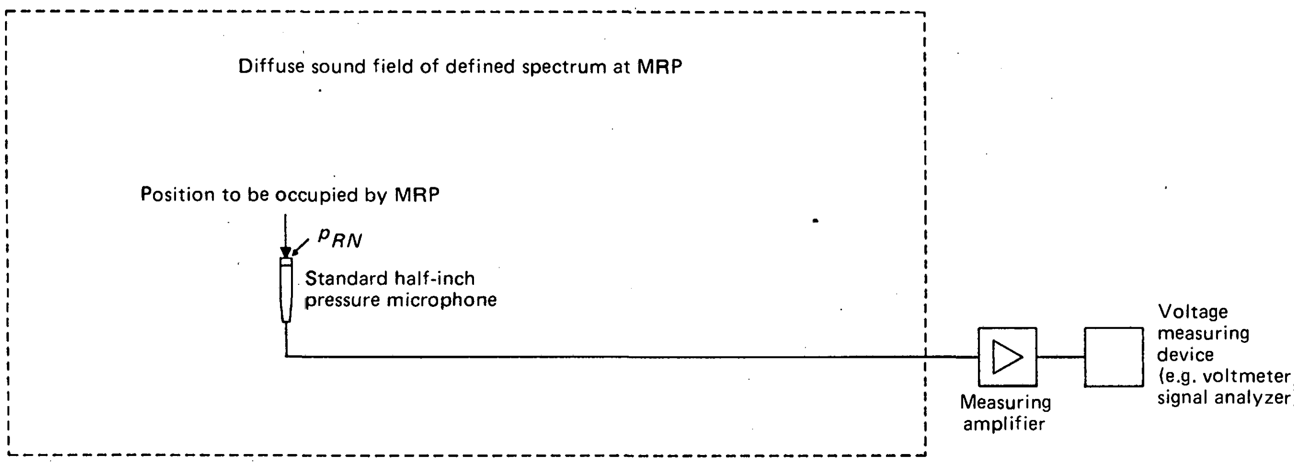


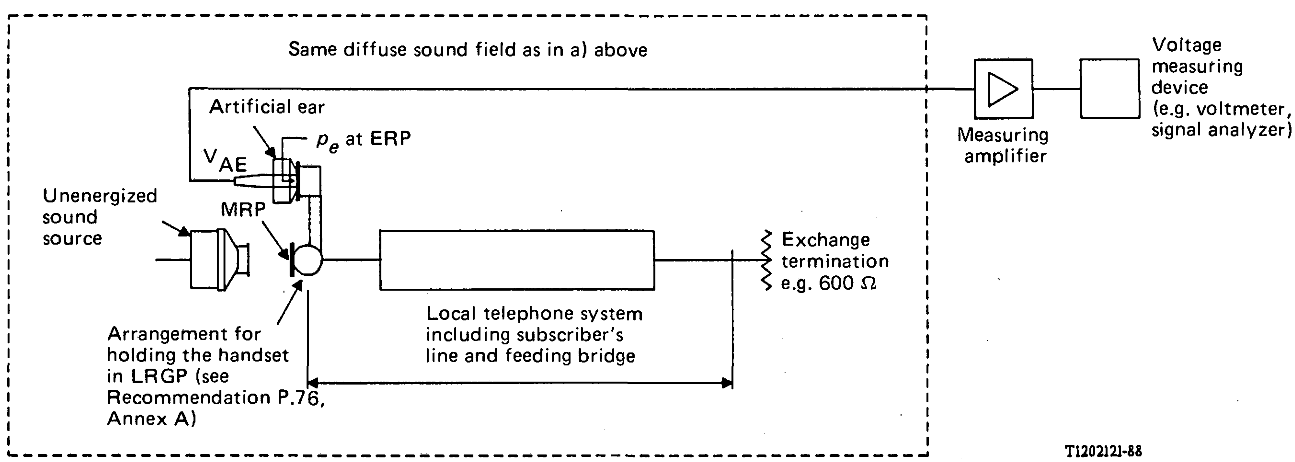
FIGURE 4/P.64

Measurement of the sidetone sensitivity of a commercial telephone system by determination of the sound pressure p_e developed in an artificial ear for a given sound signal at the MRP

The determination of the room noise sidetone sensitivity S_{RNST} is illustrated in Figure 5/P.64. For this measurement, sine wave signals are unsuitable and it is necessary to make use of continuous spectrum sound having, for example, a Hoth or pink noise spectrum (see § B.3). First, the magnitude of the diffuse field p_{RN} is determined and then the sound pressure at the artificial ear is measured.



a) Measurement of diffuse sound field at MRP



b) Measurement of p_e at the artificial ear due to diffuse sound field, p_{RN} heard via the telephone sidetone path

FIGURE 5/P.64

Using the above method, the sound pressure developed at the artificial ear usually is very low. An alternative way to determine S_{RNST} is to measure the sending sensitivity S_{mJ} using an artificial mouth and one of the methods in Annex B, using a continuous spectrum signal (e.g. §§ B.3, B.4) and then to measure the room noise sending sensitivity $S_{mJ/RN}$ using a diffuse field method such as described for room noise sidetone sensitivity above. (A detailed description of the method is given in the Handbook on Telephonometry).

The definition of Δ_{SM} is

$$\Delta_{SM} = S_{MJ/RN} - S_{MJ}$$

where S_{MJ} is the real voice sensitivity.

However, for all practical purposes, when using the artificial mouth, we may consider that Δ_{SM} is equal to Δ_{Sm}:

$$\Delta_{Sm} = S_{mJ/RN} - S_{mJ}$$

so that S_{RNST} can be determined by the approximation:

$$S_{RNST} \simeq S_{meST} + \Delta_{Sm}$$

Note 1 – For an explanation of how Δ_{Sm} may be used in the determination of Listener Sidetone Rating (LSTR) from Sidetone Masking Rating (STMR), see Recommendations P.76, P.79 and G.111.

Note 2 – In many cases, especially for carbon microphones, Δ_{Sm} , and hence also S_{RNST} is a function of the level of P_{RN}. It is recommended that in these cases the level of P_{RN} should be mentioned together with Δ_{Sm}. Typical value of P_{RN} should lie within 40-65 dBA (see Handbook on Telephonometry, § 3.3).

Note 3 – Both S_{mJ} and $S_{mJ/RN}$ should use the same techniques, e.g. wideband signals measured in ⅓ octave bands.

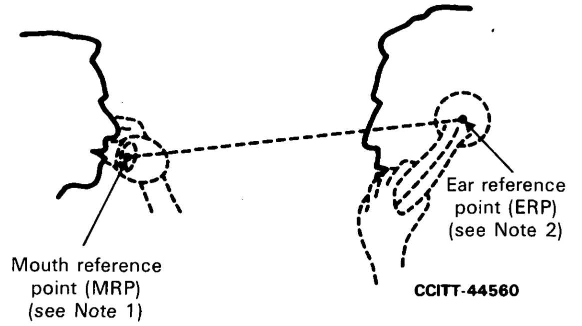
Note 4 – The approximate formulae for S_{RNST} can be deemed to be equal for linear systems.

ANNEX A

(to Recommendation P.64)

Definitions of mouth reference point and ear reference point

The definitions of mouth reference point (MRP) and ear reference point (ERP) are illustrated in Figure A-1/P.64.



Note 1 – The mouth reference point is located at a distance of 25 mm in front of the lips on the horizontal axis through the centre of the opening of the mouth. It is defined in the absence of any obstruction.

Note 2 — The ear reference point is located at the entrance to the ear canal of the listener's ear. It is defined as lying at the centre of the front plane of a circular concave earcap when sealed to the ear.

Note 3 — The ERP should not be confused with the earcap reference point (ECRP) which is a point in the earphone reference plane used as a handset reference parameter (see Recommendation P.10).

FIGURE A-1/P.64

Definitions of mouth and ear reference points

ANNEX B

(to Recommendation P.64)

Measurement of local telephone systems containing carbon microphones

For the measurement of local telephone systems containing carbon microphones, various methods have been suggested and tried. The following gives, as examples, some of these methods. These same methods can also apply to telephones using linear microphones.

Carbon microphones must be given appropriate conditioning treatment at suitable intervals during the measurement (see Recommendation P.75).

Further information can be found in [3].

Note 1 – The efficiency of the artificial mouth used is not generally constant with frequency, so it is necessary, for most of the methods described below, to insert appropriate equalization networks between the electrical signal generator and the loudspeaker of the artificial mouth. It is the free field acoustical signal which shall conform to the complex signal or the artificial voice specified.

Note 2 – It has been found that for specific applications it may be advantageous to use speech levels other than the −4.7 dBPa recommended below. This should only be done with care having due regard to the particular application. Studies carried out during the Study Period 1985-1988 have shown that better agreement with subjective test results are obtained with somewhat lower levels, e.g. over the range −4.7 to −7.0 dBPa.

B.1 The *upper envelope method* has been used in the CCITT Laboratory with success for some types of carbon microphone but has been less successful with others. The upper envelope method is as follows:

a) Determine the sensitivity as a function of frequency at the sound pressure level of −4.7 dB relative to 1 Pa. This is somewhat higher than the mean power of active speech of a talker, emitting speech at the vocal level used to determine loudness ratings in accordance with the subjective test method described in Recommendation P.78;

b) Repeat a) but with the sound pressure level increased by 10 dB;

c) Repeat a) but with the sound pressure level decreased by 10 dB;

d) Select from a), b) and c) the highest sensitivity at each frequency.

B.2 *Sweeping frequency method*

Some available types of objective instrumentation for measuring loudness-related ratings use a sweeping frequency covering the range from 200-4000-200 Hz at a periodicity of 1 sweep per second; the instantaneous level within any narrow frequency band varies as a function of frequency approximately in accordance with the spectrum of speech emitted from the human mouth.

This method should not be used for determining Δ_{SM}.

B.3 *Pink-noise method*

The handset containing the carbon microphone is placed in front of an artificial mouth producing at the MRP pink noise (power spectrum density diminishing by 3 dB/octave) over 1/3rd octave frequency bands centred on the preferred frequencies specified in ISO Standard 266-1975 at 1/3rd octave intervals in the range 100 to 8000 Hz with the band edges conforming to the filters described in IEC 225.

The total level of the signal, measured over the same bandwidth, should be −4.7 dBPa with a tolerance of ± 1.0 dB.

Note – This may not be practical with all artificial mouths, and a narrower bandwidth of 200 to 8000 Hz may have to be used for some types of artificial mouths.

The sensitivity/frequency characteristic is obtained by finding the ratio of the spectrum density of the signal delivered by the telephone system to the spectrum density of the signal obtained using a small linear microphone placed at the MRP under free-field conditions (after removing the handset).

B.4 *Shaped gaussian noise method*

The method uses shaped Gaussian noise at the MRP whose long-term average spectrum density is the same as shown in Table 1/P.50. The total level of the signal should be −4.7 dBPa ± 1 dB.

The sensitivity/frequency characteristic is obtained as in § B.3.

B.5 *Real-voice calibration*

This may be performed by measuring speech spectra emitted alternately or simultaneously from the carbon microphone under test and a calibrated linear microphone. A very small linear microphone can be mounted on the telephone being tested. Naturally the most appropriate results will be obtained when the talkers are conducting telephone conversations, but it is then difficult to have reliable knowledge of the sensitivity/frequency characteristic of the linear microphone. It is usually necessary to rely upon a suitable artificial mouth to provide the calibration of the linear microphone.

B.6 *Application of a wideband signal*

The wideband signal is generated by a pseudo-random binary sequence and is then equalized to have a long-term average spectrum density flat or as defined in Recommendation P.50. The output from the carbon microphone is then processed by fast Fourier transform (FFT) techniques. This method, like the previous method, requires calibration by a linear microphone of known sensitivity/frequency characteristic to determine the value of p_m. This method has the advantage that the frequency characteristic may be obtained with a sample of test signal of very short duration (e.g. 50 ms).

B.7 *Method using the artificial voice*

The method uses the artificial voice, having spectral and time characteristics similar to those of speech.

The sensitivity/frequency characteristic is obtained as in § B.3 above, but with the artificial mouth supplying the acoustic artificial voice, defined in Recommendation P.50.

References

[1] CCITT – Question 8/XII, Contribution COM XII-No. 1, Study Period 1989-1992.

[2] CCITT – *Handbook of Telephonometry*; ITU, Geneva, 1987.

[3] CCITT – Contribution COM XII-R 27, § B.2, Study Period 1985-1988.

Recommendation P.65

OBJECTIVE INSTRUMENTATION FOR THE DETERMINATION OF LOUDNESS RATINGS

(Malaga-Torremolinos, 1984, amended at Melbourne, 1988)

1 Introduction

This Recommendation describes the essential features of objective instrumentation suitable for the determination of loudness ratings. These features are drawn from current Recommendations relating to loudness ratings, the principles of which are defined in Recommendation P.76.

It is possible to realize objective instrumentation for loudness rating purposes in a number of ways, for example by the assembly of a number of separate instruments, each having its own defined function, and possibly under some central control, or by means of a dedicated piece of apparatus specially designed for the purpose. However, in order to ensure that loudness rating measurements made in different laboratories have an acceptable level of agreement, say ± 1 dB, it is essential that the Recommendations relating to the measurement of the electro-acoustic performance of telephone systems should be followed.

The relevant Recommendations are:

P.48 Specification for an intermediate reference system

P.51 Atificial ear and artificial mouth

P.64 Determination of sensitivity/frequency characteristics of local telephone systems to permit calculation of their loudness ratings

P.75 Standard conditioning method for handsets with carbon microphones

P.76 Determination of loudness ratings; fundamental principles

P.79 Calculation of loudness ratings [1].

2 Instrumentation

The four electro-acoustic sections that are required to be included in equipment intended for use in determining loudness ratings are described below. In each case appropriate calibration is required as a function of frequency, and calibration values recorded in the fifth section where the particular sensitivity/frequency characteristic is derived and the loudness rating calculated. If the instrumentation is to include the measurement of listener sidetone rating (LSTR), a sixth section must be provided, namely a diffuse room noise source together with appropriate facilities for calibration, measurement and analysis in one-third octave bands.

It is necessary to provide certain auxiliary apparatus, such as feeding circuits, artificial subscriber cable and exchange terminations, as required by the particular Recommendation(s) being followed for any given measurement.

2.1 *Artificial ear*

See *a)* of Figure 1/P.65.

The artificial ear in the system should be in accordance with Recommendation P.51 and contain within it a measuring amplifier so that the pressure p_e occurring at the artificial ear cavity can be measured as a function of frequency, or in frequency bands within the recording and measurement system, *e)* of Figure 1/P.65. Means must also be available to calibrate the standard microphone used in the artificial ear employing, for example, an acoustic calibrator or piston-phone.

2.2 *Artificial voice*

See *b)* of Figure 1/P.65.

An artificial mouth complying with Recommendation P.51 must be part of the system and be able to produce a prescribed sound field at the MRP 25 mm in front of the lip plane. A signal source will be part of the artificial voice and this source may be sine waves (swept or discrete frequencies) or a wideband signal (e.g. the artificial voice defined in Recommendation P.50, or shaped Gaussian noise as defined in Recommendation P.64, § B.4). Equalization and gain control should be part of the drive system to the artificial mouth such that the sound pressure at the MRP can be controlled in accordance with the requirements of Recommendation P.64, §§ B.1 and B.4, or as appropriate.

Calibration of the sound pressure and/or spectrum at the MRP may be carried out using the standard microphone used in the artificial ear of § 2.1 above, making use of the recording and measurement system of § 2.5 below to determine p_m as a function of frequency, or in frequency bands.

Mechanical means must be provided to hold the test handset in the LRGP (loudness rating guard-ring position), in accordance with the requirements of Recommendation P.76, Annex A. If handsets having carbon microphone are being tested, conditioning in accordance with Recommendation P.75 must be provided.

2.3 *Electrical termination*

See *c)* of Figure 1/P.65.

The system should contain a 600 ohm balanced terminating impedance with means for measuring the terminating voltage, V_J (see Recommendation P.64, §§ 6 and 9, as a function of frequency, or in frequency bands, using the recording and measurement system of § 2.5 below. Calibration of this section may be carried out using a calibrated voltage source.

[1] Other algorithms are being studied by Study Group XII, under Question 15/XII.

2.4 *Electrical signal source*

See *d)* of Figure 1/P.65.

An electrical signal source must be provided having a 600 ohm balanced impedance. The electrical source need not be the same as that used for the artificial voice but should either be sine waves or a wideband signal. There should be means for calibrating and adjusting the generator voltage, E_J, to the requirements of Recommendation P.64, §§ 7 and 9 over the frequency range 100-8000 Hz. This may be carried out using calibration of the electrical termination of § 2.3 above.

2.5 *Recording and measurement system*

See *e)* of Figure 1/P.65.

In order to determine the sound pressure p_e at the artificial ear or the voltage V_J at the electrical termination it will be necessary to provide a recording and measurement system. This measurement system may, using hardware or software, contain filters in order to improve signal-to-noise ratio or for analysing the output of the telephone set in 1/3rd octave frequency bands. Where a bank of 1/3rd octave filters is used these should be centred on the preferred frequencies in accordance with ISO 266 and have the characteristics in accordance with IEC Publication 225.

Within this part of the system there should be recording or storage facilities so that calibration and measurement data may be used to derive the necessary sensitivity/frequency characteristics in accordance with Recommendation P.64. The various loudness ratings are then computed in accordance with Recommendation P.79 from the sensitivity/frequency characteristics, taking into account any recognized adjustments, for example L_E or L_M. Values for L_E and L_M may be fed into the calculation using default values (e.g. those listed for L_E in Table 4/P.79) or from other more appropriate data sources when available.

2.6 *Diffuse room noise source*

See *f)* of Figure 1/P.65.

If LSTR is to be measured, a diffuse room noise source must be available, calibrated to provide a prescribed sound field at the position to be occupied by the MRP in the absence of the test head and all other obstacles, and as described in Recommendation P.64, § 9. Calibration of the diffuse sound pressure p_{RN} may be carried out using the standard microphone used in the artificial ear of § 2.1, making use of the recording and measurement system of § 2.5 to determine p_{RN} as a function of frequency in the frequency bands.

Because of the nature of room noise sidetone, it will normally be appropriate to use a diffuse sound pressure p_{RN} that is much lower than the value of -4.7 dBPa used for p_m in determining STMR and SLR. Typical values for p_{RN} would lie in the range 40-65 dB SPL (-54 to -29 dBPa, A weighted), and it should have a frequency spectrum appropriate for the application, for example as given in Supplement No. 13, § 2. The actual level and type of noise should always be stated in quoting test results.

3 Measurements

Facilities should be provided to enable the various sections of the instrumentation to be connected allowing the measurement of the necessary sensitivity/frequency characteristics and calculation of the loudness ratings.

A summary of these interconnections, together with the sensitivity/frequency characteristics (SFC) measured for particular loudness rating determinations, are given below.

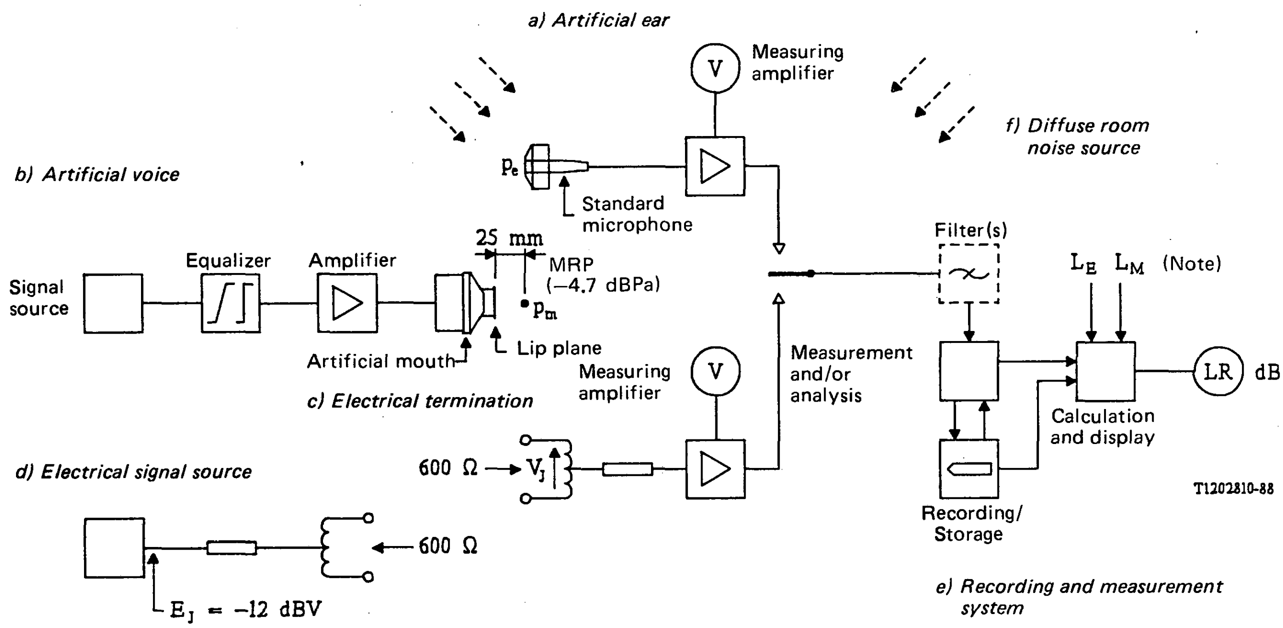
3.1 *Send loudness rating (SLR)*

Source: *b)* of Figure 1/P.65

Load: *c)* of Figure 1/P.65

Send SFC given by:

$$S_{mJ} = 20 \log_{10} \frac{V_J}{p_m} \text{ dB}$$



Note – The artificial real mouth corrections L_M has been assumed to be zero but is currently under study in Question 8/XII.

FIGURE 1/P.65

Essential features of instrumentation for the determination of loudness ratings

3.2 *Receive loudness rating (RLR)*

Source: *d)* of Figure 1/P.65

Load: *a)* of Figure 1/P.65

Receive SFC given by:

$$S_{Je} = 20 \log_{10} \frac{p_e}{\frac{1}{2} E_J} \text{ dB}$$

3.3 *Sidetone masking rating (STMR)* (Talker Sidetone)

Source: *b)* of Figure 1/P.65

Load: *a)* of Figure 1/P.65

Sidetone SFC is given by:

$$S_{meST} = 20 \log_{10} \frac{p_e}{p_m} \text{ dB}$$

Note – The quantity L_{meST} used in the calculation of STMR is given by:

$$L_{meST} = -S_{meST} \text{ dB}$$

3.4 *Listener sidetone rating (LSTR)*

Source: *f)* of Figure 1/P.65

Load: *a)* of Figure 1/P.65

Room noise sidetone SFC is given by:

$$S_{RNST} = 20 \log_{10} \frac{p_e}{p_{RN}} \text{ dB}$$

3.5 *Overall loudness rating (OLR) (Overall Send + Receive (OSR))*

Source: *b)* of Figure 1/P.65

Load: *a)* of Figure 1/P.65

Overall SFC given by:

$$S_{me} = 20 \log_{10} \frac{p_e}{p_m} \text{ dB}$$

3.6 *JLR Junction loudness rating*

Source: *d)* of Figure 1/P.65

Load: *c)* of Figure 1/P.65

Junction loss/frequency characteristics given by:

$$X_{JJ} = 20 \log_{10} \frac{\frac{1}{2} E_J}{V_J} \text{ dB}$$

Note – Impedance terminations of 600 ohms are assumed.

Recommendation P.66

METHODS FOR EVALUATING THE TRANSMISSION PERFORMANCE OF DIGITAL TELEPHONE SETS [1)]

(Melbourne, 1988)

1 Introduction

The CCITT recommends the following method to evaluate the voice transmission performance of a digital telephone set using encoding conforming to Recommendation G.711 (see also Recommendation P.31). A digital telephone set is one in which the A/D and D/A converters are built in and the connection to the network is via a digital bit-stream. This poses a fundamental problem in applying existing performance and measurement techniques, such as Recommendations P.64, P.34 and P.38, since these are generally given in terms of a transfer function of analogue input and output quantities, e.g. a frequency response. The principles involved in this Recommendation are applicable to handset, headset and hands-free operation; however, at present only procedures applicable to handset operation have been developed.

2 Approaches for testing digital telephones

There are two methods for evaluating the transmission performance of a digital telephone, the codec approach and the direct approach.

In the short term, use of the codec approach is advocated since many Administrations already have some experience with this methodology (see Recommendation O.133).

1) The specifications in this Recommendation are subject to future enhancement and therefore should be regarded as provisional.

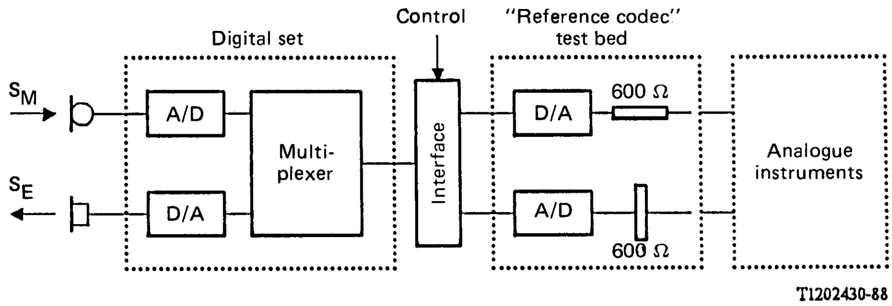


FIGURE 1/P.66

Digital telephone test arrangement (codec approach)

2.1 *Codec approach*

In this approach, shown in Figure 1/P.66, a codec is used to convert the companded digital input/output bit-stream of the telephone set to the equivalent analogue values, so that existing test procedures and equipment can be used. This codec should be a high-quality codec whose characteristics are as close as possible to ideal (see § 5).

2.2 *Direct digital processing approach*

In this approach, shown in Figure 2/P.66, the companded digital input/output bit-stream of the telephone set is operated upon directly.

Note – This approach is still under study in Question 38/XII.

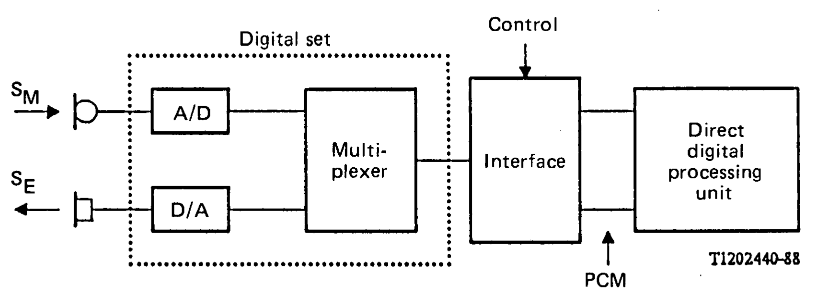


FIGURE 2/P.66

Digital telephone test arrangement (direct digital processing approach)

3 Definition of 0 dB reference point

To preserve compatibility with existing codecs already in use in local digital switches, which are defined as a 0 dBr point, the codec (A- or μ-law) sould be defined as follows:

- D/A converter: a digital test sequence (DTS) representing the PCM equivalent of an analogue sinusoidal signal whose r.m.s. value is 3.14 dB (A-law) or 3.17 dB (μ-law) below the maximum full-loaded capacity of the codec will generate 0 dBm across a 600 ohm load;
- A/D converter: a 0 dBm signal generated from a 600 ohm source will give the digital test sequence (DTS) representing the PCM equivalent of an analogue sinusoidal signal whose r.m.s. value is 3.14 dB (A-law) or 3.17 dB (μ-law) below the maximum full load capacity of the codec;

  where DTS is defined as a periodic sequence of character signals as given in Tables 5/G.711 and 6/G.711.

4 Definition of interfaces

The digital telephone test equipment will, in general, be connected to the telephone under test through an interface.

Such an interface should be able to provide all the signalling and supervisory sequences necessary for the telephone set to be working in all test modes. The interface must be capable of converting the digital output stream from the tested set (which may be in various formats, depending on the specific type of telephone set, e.g. conforming to Recommendation I.412 for ISDN sets), to a form compatible with the test equipment. Interfaces can be applied for sending and receiving separately, taking into account telephone sets which are connected to various types of exchanges.

5 Codec specification

5.1 *Ideal codec*

The ideal codec consists of an independent encoder and decoder whose characteristics are hypothetical and comply with Recommendation G.711. The ideal encoder is a perfect analogue-to-digital converter preceded by an ideal low-pass filter (assumed to have no attenuation/frequency distortion and no envelope-delay distortion), and may be simulated by a digital processor.[2] The ideal decoder is a perfect digital-to-analogue converter followed by an ideal low-pass filter (assumed to have no attenuation/frequency distortion and no envelope-delay distortion), and which may be simulated by a digital processor[2].

For the measurement of the sending side of a telephone set, the output digital signal is converted by the decoder to an analogue signal. The electrical characteristics of this output signal are measured using conventional analogue instruments. For the measurement of the receiving side of a telephone set, the analogue output from a signal source is converted to a digital signal by the ideal encoder and fed to the receiving input of the digital telephone set.

5.2 *Reference codec*

A practical implementation of an ideal codec may be called a reference codec (see Recommendation O.133, § 4).

For the reference codec, characteristics such as attenuation/frequency distortion, idle channel noise, quantizing distortion, etc. should be better than the requirements specified in Recommendation G.714, so as not to mask the corresponding parameters of the set under test. A suitable reference codec may be realized by using:

1) at least 14 bit linear A/D and D/A converters of high quality, and transcoding the output signal to the A- or μ-law PCM format;
2) a filter response that meets the requirements of Figure 3/P.66.

5.2.1 *Analogue interface*

The output and input impedances return loss and longitudinal conversion losses of the analogue interface of the reference codec should be in accordance with Recommendation O.133, § 3.1.1.

5.2.2 *Digital interface*

The fundamental requirements for the reference codec digital interface are given in the appropriate Recommendations (e.g. I.430-Series Recommendations for ISDN telephone sets).

6 Measurement of digital telephone transmission characteristics

Use of the codec test approach means that test procedures for digital telephone sets in general follow those for analogue sets (see Recommendation P.64). The reference codec should meet the requirements of § 5. An important difference, however, concerns the test circuits themselves, see Figures 4/P.66 to 10/P.66.

The set is connected to the interface and is placed in the active call state.

Note – When measuring digital telephone sets, it is advisable to avoid measuring at sub-multiples of the sampling frequency. There is a tolerance on the frequencies of ± 2% which may be used to avoid this problem, except for 4 kHz where only the −2% tolerance may be used.

---

[2] This characteristic can be realized, for example, using oversampling techniques and digital filters.

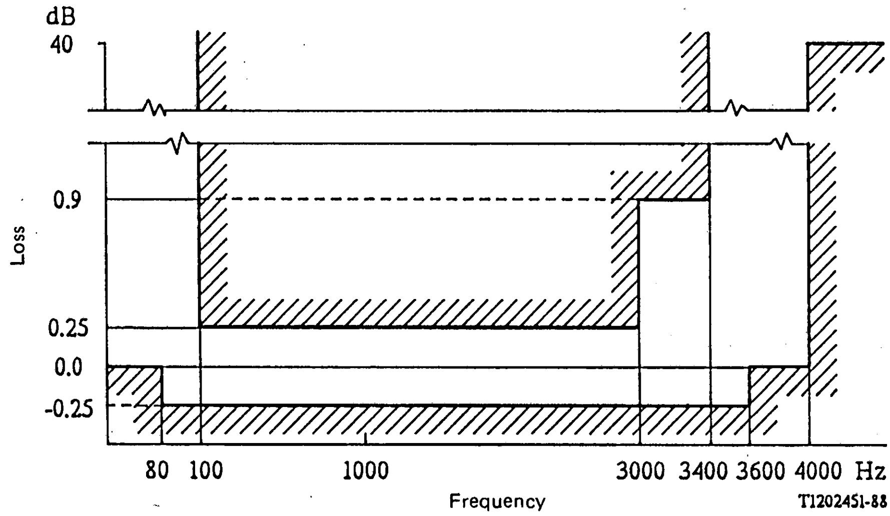


FIGURE 3/P.66

Attenuation/frequency distortion of the sending or receiving sides of the reference codec

6.1 *Sending*

6.1.1 *Sending frequency characteristic*

The sending frequency characteristic is measured according to Recommendation P.64 using the measurement set-up shown in Figures 4/P.66 or 5/P.66, depending on the excitation signal used.

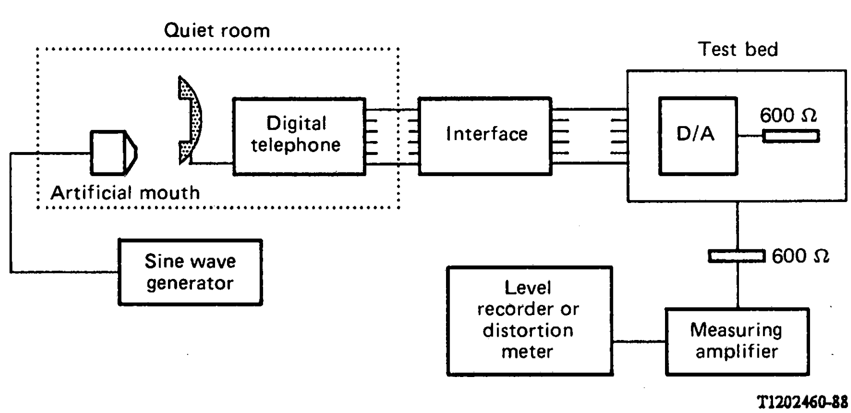


FIGURE 4/P.66

Measurements of sending frequency characteristic swept sine wave technique

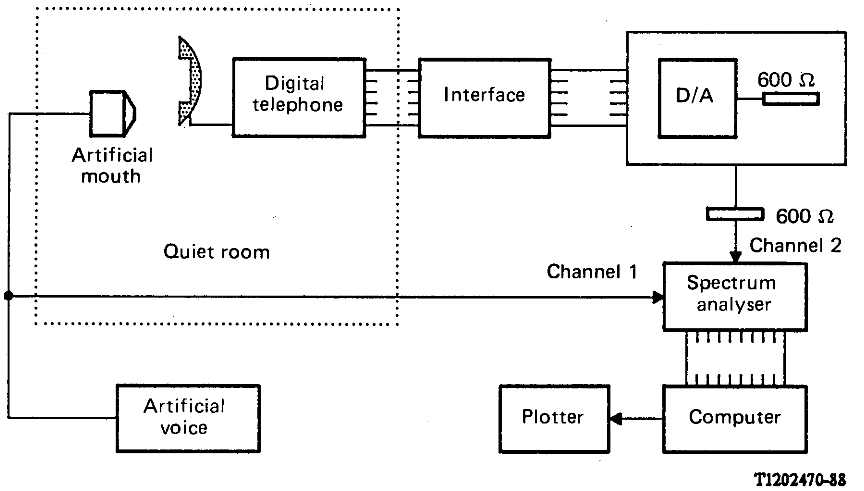


FIGURE 5/P.66

Measurement of sending frequency characteristic – wideband technique

6.1.2 *Send loudness rating*

This should be calculated from the sensitivity/frequency characteristic determined in § 6.1.1 by means of Recommendation P.79.

Note – Other methods for calculating loudness ratings used by some Administrations for their own internal planning purposes can be found in Supplement No. 19.

6.1.3 *Distortion*

Under study in Question 38/XII, but a method used by several European Administrations can be found in [1].

6.1.4 *Noise*

Under study in Question 38/XII, but a method used by several European Administrations can be found in [1].

6.1.5 *Linearity*

Under study in Question 38/XII, but a method used by several European Administrations can be found in [1].

6.1.6 *Discrimination against out-of-band input signal*

Under study in Question 38/XII, but a method used by several European Administrations can be found in [1].

6.2 *Receiving*

6.2.1 *Receiving frequency characteristic*

The receiving frequency characteristic is measured according to Recommendation P.64 using the measurement set-up shown in Figures 6/P.66 or 7/P.66, depending on the excitation signal used.

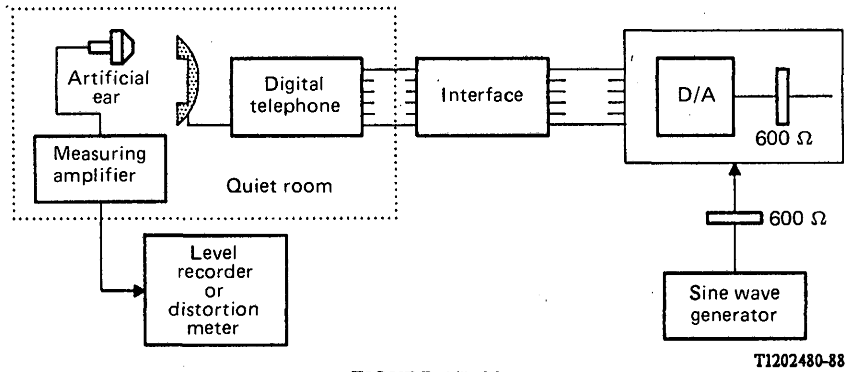


FIGURE 6/P.66

Measurement of receiving frequency characteristic – swept sine wave technique

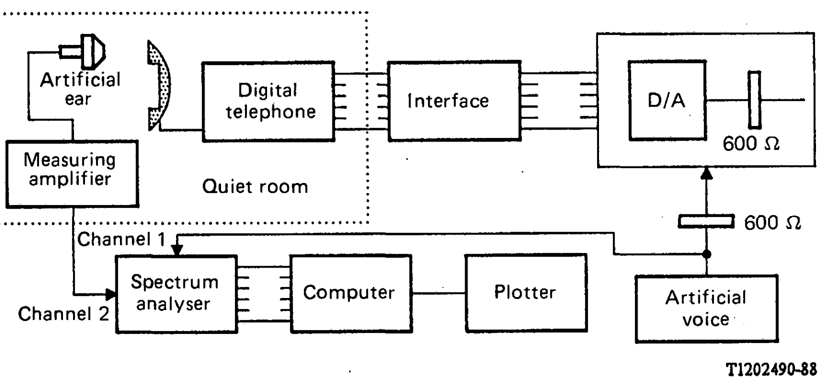


FIGURE 7/P.66

Measurement of receiving frequency characteristic – wideband technique

6.2.2 *Receiving loudness rating*

This should be calculated from the sensitivity/frequency characteristic determined in § 6.2.1 by means of Recommendation P.79.

Note – Other methods for calculating loudness rating used by some Administrations for their own internal planning purposes can be found in Supplement No. 19.

6.2.3 *Distortion*

Under study in Question 38/XII, but a method used by several European Administrations can be found in [1].

6.2.4 *Noise*

Under study in Question 38/XII, but a method used by several European Administrations can be found in [1].

6.2.5 *Linearity*

Under study in Question 38/XII, but a method used by several European Administrations can be found in [1].

6.2.6 *Supurious out-of-band signals*

Under study in Question 38/XII, but a method used by several European Administrations can be found in [1].

6.3 *Sidetone*

Provision should be made for driving the microphone of the telephone set under test as described in § 6.1 and measuring the receiver output as described in § 6.2. The recommended method of measuring sidetone is with the microphone and receiver mounted in the same handset, and using a test fixture which includes the artificial mouth and the artificial ear located relative to each other in accordance with Recommendation P.64.

Note – Care should be taken to avoid mechanical coupling between the artificial mouth and the artificial ear.

6.3.1 *Sidetone frequency characteristic*

6.3.1.1 *Talker sidetone frequency characteristic*

The talker sidetone frequency characteristic is measured according to Recommendation P.64 using the measurement set-up of Figures 8/P.66 or 9/P.66 depending on the excitation signal used.

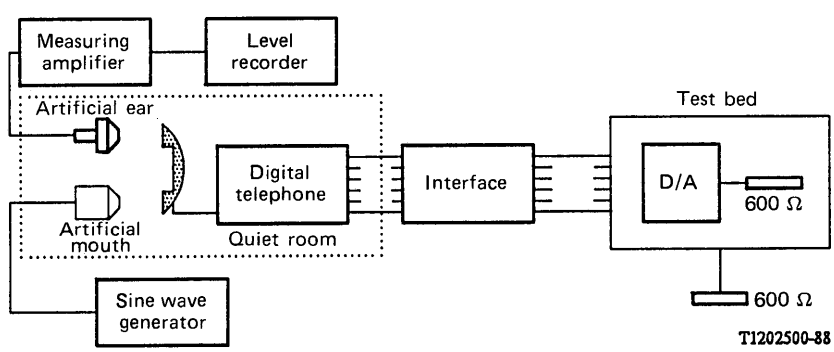


FIGURE 8/P.66

Measurement of talker sidetone frequency characteristic – swept sine wave technique

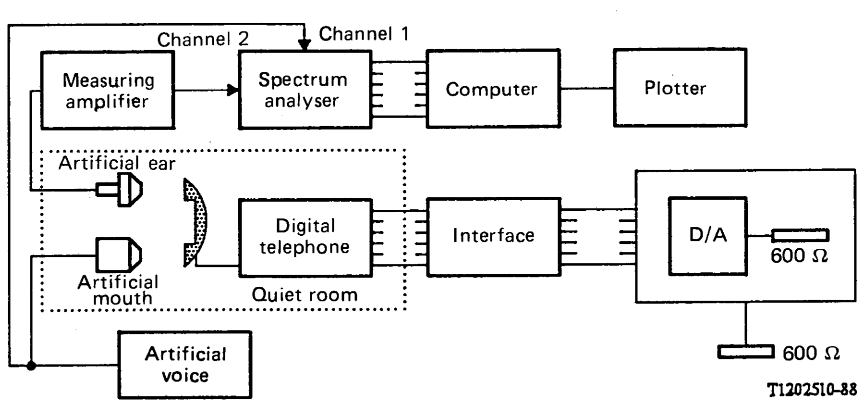


FIGURE 9/P.66

Measurement of talker sidetone frequency characteristic – wideband technique

6.3.1.2 *Listener sidetone frequency characteristic*

The listener sidetone frequency characteristic is measured according to Recommendation P.64 using the measurement set-up of Figure 10/P.66.

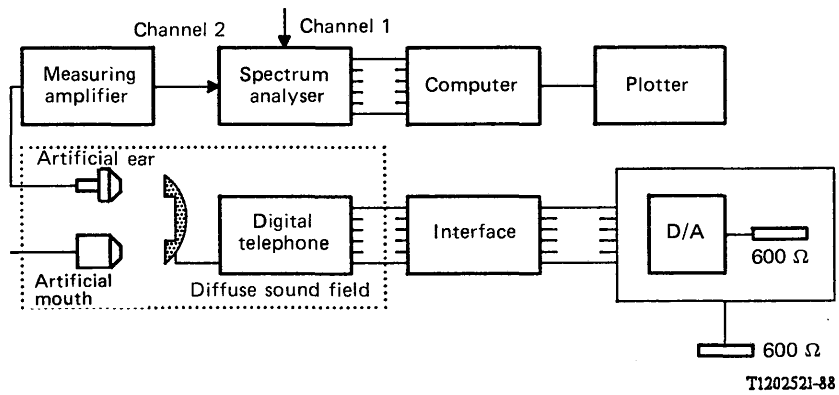


FIGURE 10/P.66

Measurement of listener sidetone frequency characteristic

6.3.2 *Sidetone masking rating*

This should be calculated from the sensitivity/frequency characteristic determined in § 6.3.1.2 by means of Recommendation P.79.

6.3.3 *Listener sidetone rating*

This should be calculated from the sensitivity/frequency characteristic determined in § 6.3.1.2 by means of Recommendation P.79.

6.4 *Echo return loss*

Under study in Question 38/XII, but a method used by several European Administrations can be found in [1].

6.5 *Delay*

Under study in Question 38/XII, but a method used by several European Administrations can be found in [1].

References

[1] CCITT – Contribution COM XII-No.179, *Transmission aspects for digital telephony* (Norway), Study Period 1985-1988.

SECTION 6

MEASUREMENTS RELATED TO SPEECH LOUDNESS

Recommendation P.75

STANDARD CONDITIONING METHOD FOR HANDSETS WITH CARBON MICROPHONES

(Geneva, 1972; amended at Malaga-Torremolinos, 1984, Melbourne, 1988)

1 Since the characteristics of carbon microphones are strongly dependent on conditioning techniques, it is necessary to follow a consistent procedure prior to measuring sensitivity/frequency characteristics in order to obtain reproducible results. The CCITT recommends that for best reproducibility, automatic mechanical conditioning be used. The following steps are specified for the *standard conditioning method*:

a) Place the handset in a holding fixture with the handset clamped in a position corresponding to that in which the microphone is going to be measured [e.g. loudness rating guard-ring position (LRGP) according to Annex A of Recommendation P.76].

b) Connect the microphone or telephone set terminals as required to the d.c. feed circuit and appropriate terminating loading.

c) Turn the feed current on. After 5 seconds, condition the microphone by rotating it smoothly. Rotation is made such that the plane of the granule bed moves through an arc of at least 180°. The procedure is repeated twice with the handset coming to rest finally in the test position. The time of each rotation cycle should lie within the range of 2 to 12 seconds.

2 When carrying out subjective tests with a carbon microphone telephone set, the conditioning of the handset should be done by the talker. This conditioning should conform to the conditioning for objective measuring as described under § 1 above insofar as it is practicable.

DETERMINATION OF LOUDNESS RATINGS; FUNDAMENTAL PRINCIPLES

(Geneva, 1976; amended at Geneva, 1980 Malaga-Torremolinos, 1984, Melbourne, 1988)

Preface

This Recommendation is one of a set of closely related Recommendations concerned with determination of loudness ratings. The present one deals with the fundamental principles and the others, as follows, deal with certain additional matters [1)].

Recommendation P.48	Specification for an intermediate reference system
Recommendation P.78	Subjective testing method for determination of loudness ratings in accordance with Recommendation P.76
Recommendation P.64	Determination of sensitivity/frequency characteristics of local telephone systems to permit calculation of their loudness ratings
Recommendation P.79	Calculation of loudness ratings
Recommendation P.65	Objective instrumentation for the determination of loudness ratings

1 Introduction

A speech path is, broadly, a transmission path that exists between a talker's mouth and the ear of a listener or, in the case of sidetone, between the mouth and ear of a talker. In typical face-to-face conversation, the speech is transmitted by means of the air path connecting the mouth and ear. Depending on environmental conditions, transmission may be:

a) more or less direct, as in the case of two persons conversing in an open, unobstructed location, such as a golf course;

b) largely indirect, as in the case of two persons conversing in a small, hard surfaced room where a large proportion of the energy reaching the ear may be due to reflections from the walls, ceilings and floor; or

c) something between the two extremes of *a)* and *b)*.

In the case of telephony, the air path is replaced by a system comprising:

a) an air path from the mouth to the telephone microphone;

b) an air path between the telephone earphone and the ear; and

c) a telephone connection consisting of the microphone, earphone and interconnecting circuitry together with a similar system for the reverse direction of transmission. The two situations – face-to-face and using the telephone – differ appreciably in detail but, for speech transmission purposes, they are alike insofar as their function is to provide a means of both-way speech communication.

Telephone engineering is concerned with providing telephone connections which, while not identical to the face-to-face situation, are comparable in effectiveness for providing a means of exchanging information by speech; such telephone connections should also optimize customer satisfaction within technical and economic constraints.

1) The present Recommendation together with Recommendations P.48, P.78 and P.79 provide complete definitions of overall, sending, receiving and junction loudness ratings, and Administrations are invited to use them to further their studies of Question 19/XII [1].

Various tools are used by transmission engineers in planning, design and assessment of the performance of telephone networks. Reference equivalent, based on the criterion of loudness of speech emitted by the talker and perceived by the listener, has been one of the most important of these tools; it provides a measure of the transmission loss, from mouth to ear, of a speech path.

The *reference equivalent method* is defined in Recommendations P.42 and P.72 *Red Book* and its fundamental principles are briefly explained in [2]. The method for determining *loudness ratings* of local telephone circuits is based upon rather similar fundamental principles but comprises modifications which render it much more flexible and should greatly simplify transmission planning.

A desire to depart from use of reference equivalents as defined by Recommendation P.72 *Red Book* arises from the following reasons:

1) reference equivalents cannot be added algebraically; discrepancies of at least ± 3 dB are found;

2) replication accuracy of reference equivalents is not good; changes in crew can cause changes of as much as 5 dB;

3) increments of real (distortionless) transmission loss are not reflected by equal increments of reference equivalent; 10 dB increase in loss results in an increase in reference equivalent of only about 8 dB.

Use of loudness ratings defined in accordance with the principles given below should largely obviate these difficulties.

In addition to these advantages, the same values of loudness ratings should be obtained whether the determination is by subjective tests, by calculation based on sensitivity/frequency characteristics or by objective instrumentation. The fundamental principles of the method are described below and these differ from those applicable to reference equivalents by the least possible extent to achieve the desirable flexibility.

The loudness rating (which has the dimensions and sign of "loss") is, in principle, like the reference equivalent, defined by the amount of loss inserted in a reference system to secure equality of perceived loudness to that obtained over the speech path being measured. Practical telephone connections are composed of several parts connected together. To enable the transmission engineer to deal with these parts in different combinations, loudness ratings must be defined in a suitable manner so that "overall", "sending", "receiving" and "junction" ratings can be used.

"Sidetone" loudness ratings can also be determined in an analogous manner. Sidetone reference equivalent is defined in Recommendation P.73 *Red Book* and sidetone loudness ratings are defined in § 3 below.

2 Definitions of loudness ratings for principal speech paths

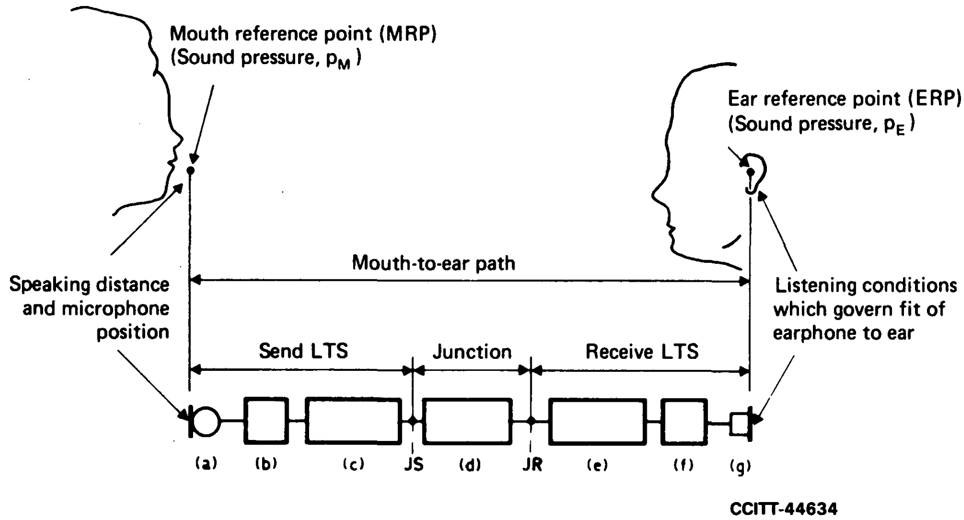
2.1 *General*

§ 2 deals with principal speech paths, namely from a talker at one end of a connection to a listener at the other. Sidetone paths are treated in § 3 below.

In general, loudness ratings are not expressed directly in terms of actual perceived loudness but are expressed in terms of the amounts of transmission loss, independent of frequency, that must be introduced into an *intermediate* reference speech path and the *unknown* speech path to secure the same loudness of received speech as that defined by a fixed setting of NOSFER. This implies that some interface exists or could, by some arrangement, be found in the unknown speech path into which the transmission loss can be introduced. In practice the unknown speech path is composed of a sending local telephone circuit coupled to a receiving local telephone circuit through a chain of circuits interconnecting the two local systems[2]. Figure 1/P.76 shows this subdivision of one principal speech path of a telephone connection. The interfaces JS and JR separate the three parts of the connection to which loudness ratings are assigned, namely: *sending loudness rating*, from the mouth reference point to JS; *receiving loudness rating* from JR to the ear reference point; and *junction loudness rating* from JS to JR. The *overall loudness rating* is assigned to the whole speech path from mouth reference point to ear reference point.

---

[2] See Annex B for explanation of certain terms.



Note – (a) represents the microphone of the sending local telephone system;

(b) represents the electrical circuit of the telephone set of the sending local telephone system;

(c) represents the subscriber's line and feeding/transmission bridge of the sending local telephone system;

(d) represents the chain of circuits interconnecting the two local systems;

(e) represents the subscriber's line and feeding/transmission bridge of the receiving local telephone system;

(f) represents the electrical circuit of the telephone set of the receiving local telephone system;

(g) represents the earphone of the receiving local telephone system.

FIGURE 1/P.76

Subdivision of a telephone connection

Note that in practical telephone connections:

a) the transmission loss of the junction may be frequency dependent;

b) the image impedances of the "junction" may not be constant with frequency and may not be resistive;

c) the impedances of the local telephone systems presented to the junction at JS and JR may not be constant with frequency and may not be resistive;

d) impedance mismatches may be present at JS or JR or both.

Overall loudness ratings (OLRs), sending loudness ratings (SLRs), receiving loudness ratings (RLRs) and junction loudness ratings (JLRs) are defined so that the following equality is achieved with sufficient accuracy for practical telephone connections.

$$\text{OLR} = \text{SLR} + \text{RLR} + \text{JLR}$$

2.2 *Definitions of overall, sending, receiving and junction loudness ratings*

Figure 2/P.76 shows the principles used to define the overall, sending, receiving and junction loudness ratings.

2.2.1 *Overall loudness rating*

Path 1 in Figure 2/P.76 shows the complete unknown speech path subdivided into local telephone systems and junction. In this example the junction comprises a chain of circuits represented by trunk junctions (JS-NS and NR-JR) and trunk circuits (NS-IS, IS-IR and IR-NR). A suitable arrangement for inserting transmission loss independent of frequency must be provided at some point such as in IS-IR.

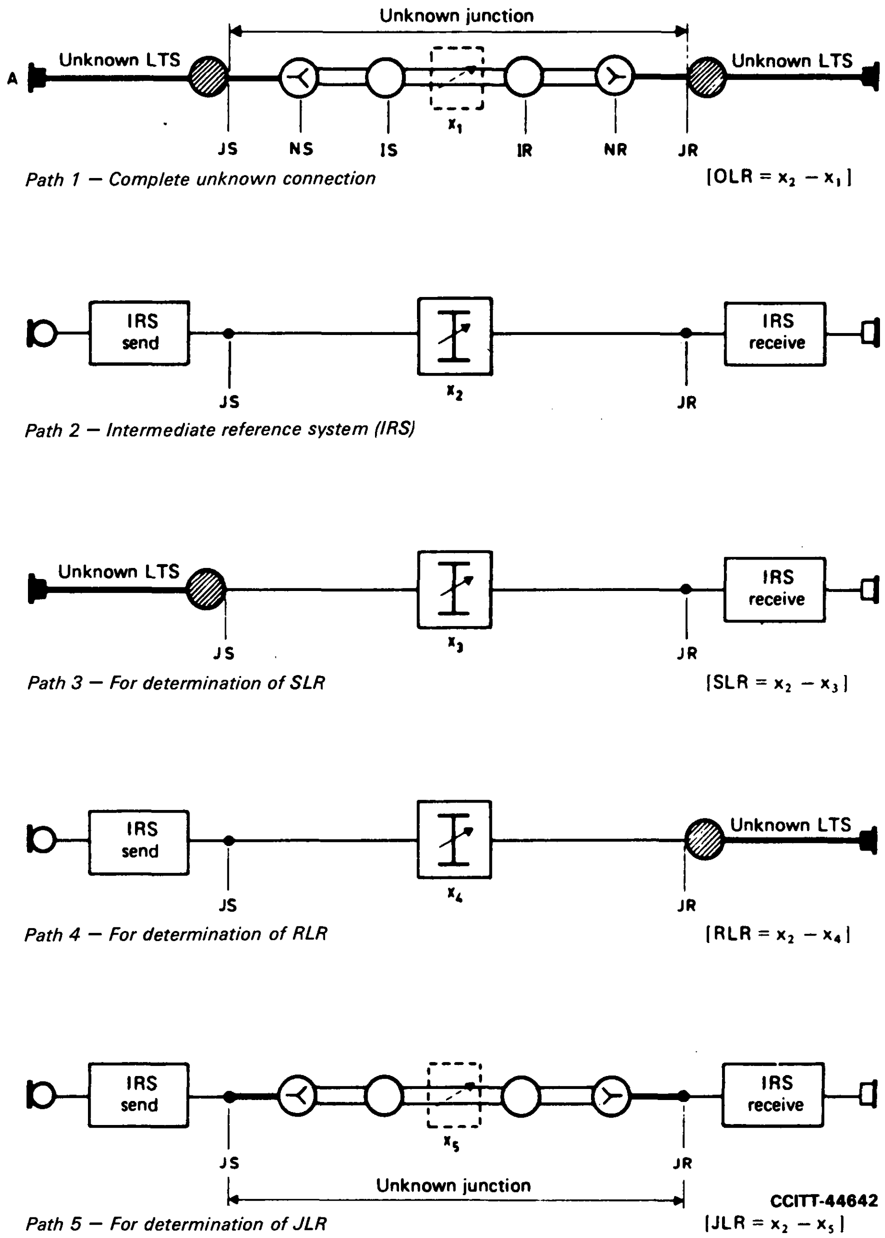


FIGURE 2/P.76

Principles used for defining OLR, SLR, RLR and JLR

Path 2 shows the complete intermediate reference system (IRS) with its adjustable, non-reactive, 600 ohms junction between JS and JR.

The level of received speech sounds to which the additional loss x_1 in Path 1 and the junction attenuator setting x_2 of Path 2 are both adjusted is defined by using the fundamental reference system NOSFER with its attenuator set at 25 dB. When these adjustments have been made, the overall loudness rating (OLR) of the complete unknown connection is given by $(x_2 - x_1)$ dB.

2.2.2 *Sending loudness rating*

Path 3 in Figure 2/P.76 shows the IRS with its sending part replaced by the local telephone system of the unknown. The junction is adjusted to produce, via Path 3, the same loudness of received speech sounds as the NOSFER with its attenuator set at 25 dB. If x_3 is the required setting in Path 3, the sending loudness rating (SLR) is given by $(x_2 - x_3)$ dB.

2.2.3 *Receiving loudness rating*

Path 4 in Figure 2/P.76 shows the IRS with its receiving part replaced by the local telephone system of the unknown.

The junction is adjusted to produce via Path 4 the same loudness of received speech sounds as the NOSFER with its attenuator set at 25 dB. If x_4 is the required setting in Path 4, the receiving loudness rating (RLR) is given by $(x_2 - x_4)$ dB.

2.2.4 *Junction loudness rating*

Path 5 in Figure 2/P.76 shows the IRS with its junction replaced by the unknown chain of circuits as located in Path 1 of Figure 2/P.76 between JS and JR. The arrangement for introducing transmission loss, independent of frequency, must be provided as was required in Path 1. The additional loss is adjusted to produce, via Path 5, the same loudness of received speech as the NOSFER with its attenuator set at 25 dB. If x_5 is the required additional loss in Path 5, the junction loudness rating is given by $(x_2 - x_5)$ dB.

2.3 *Conditions under which loudness ratings are determined*

2.3.1 *General*

The loudness of received speech sounds depends upon certain factors that are not well defined under practical conditions of use, but must be defined as precisely as possible to obtain accurately reproducible loudness ratings. Clearly, as shown in Figure 1/P.76, the loudness rating is largely governed by the characteristics of the mouth-to-ear path. This path can be made precise by defining a *mouth reference point* at which the sound pressure p_M of speech emitted by the talker is measured or referred, and an ear reference point at which to measure or to which to refer the sound pressure p_E of speech reproduced by the earphone. These points can be chosen in a fairly arbitrary manner and this becomes important when loudness ratings are to be determined objectively; suitable definitions for such purposes are given in Recommendation P.64 which deals with measurement of sending and receiving sensitivity/frequency characteristics.

It is essential, however, to define vocal level, speaking distance, microphone position and listening conditions which govern the fit of the earphone to the ear. These are indicated in Figure 1/P.76. The essential features that define the conditions under which loudness ratings are determined are indicated in Table 1/P.76.

Some remarks on the items listed in Table 1/P.76 are given below.

2.3.2 *Intermediate reference system*

The intermediate reference system is defined in Recommendation P.48. It has been chosen with the following in mind:

a) It shall correspond approximately, as far as the shapes of sending and receiving frequency characteristics are concerned, with those of national sending and receiving systems in use at present and likely to be used in the near future. For this reason the frequency bandwidths for sending and receiving parts are confined to the nominal range 300-3400 Hz[3)].

b) The absolute sensitivity has been chosen to reduce as much as possible changes in values from reference equivalents to loudness ratings.

c) In external form its handsets are similar to conventional handsets used in actual telephone connections.

[3)] The IRS is specified for the range 100-5000 Hz (see Recommendation P.48). The nominal range 300-3400 Hz specified is intended to be consistent with the nominal 4 kHz spacing of FDM systems, and should not be interpreted as restricting improvements in transmission quality which might be obtained by extending the transmitted frequency bandwidth.

TABLE 1/P.76

Conditions under which loudness ratings are determined

No.	Item specified	Specification
1	Intermediate reference system	Recommendation P.48
2	Vocal level of speaker	As Recommendation P.72 (Red Book)
3	Level of received speech sounds at which loudness is judged constant	NOSFER set at 25 dB
4	Handset position relative to talker's mouth	See Annex A
5	Direction of speech	Head erect
6	Handset arrangement for listening	See § 2.3.7
7	Conditioning of carbon microphones	Recommendation P.75

2.3.3 *Vocal level of speaker*

The vocal level at which speech is emitted from the speaker's mouth conforms to that in use for determining reference equivalents and is defined in Recommendation P.72 *Red Book.* This approximates the level actually used by customers under good transmission conditions. It is defined in terms of the speech level at the output of the NOSFER sending system.

2.3.4 *Listening level*

The level of received speech sounds at which loudness is judged constant is defined by the vocal level (see § 2.3.3 above) and the setting (25 dB) of NOSFER against which all the speech paths shown in Figure 2/P.76 are adjusted. This corresponds to a fairly comfortable listening level of the same order as that commonly experienced by telephone users.

2.3.5 *Handset position*

The position of the telephone handset relative to the talker's mouth is defined in Annex A to this Recommendation. It is intended to approximate fairly well the position used by customers under real telephone connections. The definition covers not only the distance between lips and mouthpiece but also the attitude of the microphone relative to the horizontal axis through the centre of the lips. It is defined in such a way that the lips-to-mouthpiece distance becomes greater as the length of a handset is increased.

2.3.6 *Direction of speech*

The speaker shall hold his head erect and it will be assumed that speech is emitted horizontally from his mouth.

2.3.7 *Handset arrangement for listening*

The listener shall hold the handset in his hand with the earphone placed comfortably against his ear.

2.3.8 *Conditioning of carbon microphones*

Telephone handsets with carbon microphones usually require to be conditioned. This shall be done in accordance with Recommendation P.75.

3 Sidetone loudness ratings

It is necessary to examine the effects of telephone sidetone on the subscriber when considered both as a talker and as a listener. In each case, studies have shown that control of the higher frequencies (> 1000 Hz) in the telephone sidetone path is important to preserve good conversational conditions in high-level room noise and/or on long-line connections. Sidetone loudness rating methods that place more weight on these higher frequencies are therefore required; suitable methods are described below.

3.1 *Talker Sidetone*

3.1.1 *Definition of sidetone masking rating (STMR)*

When a telephone subscriber speaks, his own voice reaches his ear by several paths (see Figure 3/P.76):

a) through the telephone set circuit from microphone to earphone due to mismatch of the hybrid balance impedance within the set and the line impedance;

b) through the mechanical path within the human head;

c) through the acoustic path to the ear and involving leakage at the earcap and human ear interface;

d) through the mechanical path along a handset handle [although this may be measured in fact as a contribution to a) above].

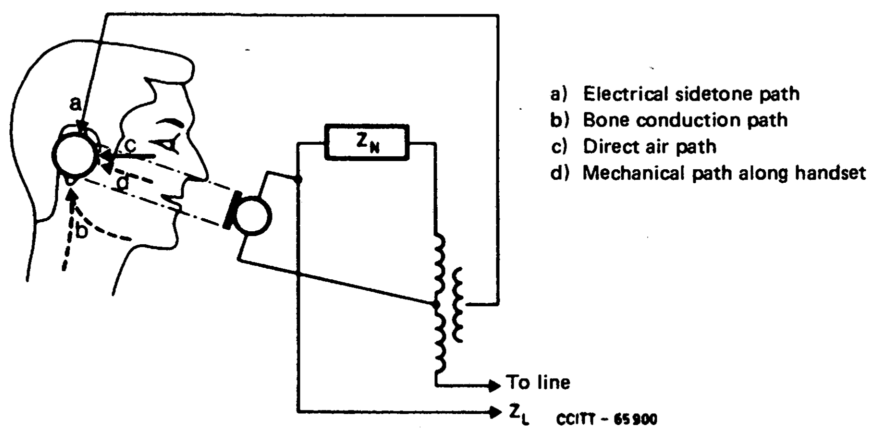


FIGURE 3/P.76

Sidetone paths through which a telephone subscriber may hear his own voice

Determination of these sidetone paths will usually resolve into two main measurements, a) + d) and b) + c). Each is referred to the speech signal at the mouth reference point (MRP) and the measurement made at the ear reference point (ERP).

Thus L_{MEST} is the loss from the mouth to ear (MRP to ERP) of the telephone sidetone path, and L_{MEHS} is the loss from mouth to ear (MRP to ERP) of the human sidetone path.

Note – Recommendation P.64, § 8 describes a method for the measurement of S_{meST}, the sidetone sensitivity/frequency characteristic of a telephone set using the artificial mouth and ear, from which an estimate of S_{MEST} using the human mouth and ear may be obtained by adding correction L_M and L_E as explained in the text. Thus:

$$L_{MEST} = -S_{MEST} \text{ in dB}$$

L_{MEST} and L_{MEHS} are each usually measured at a number of frequencies in the ISO range of ⅓rd octave frequencies, typically at least 200 to 4000 Hz. Where complex signals are used (for example, during the measurement of L_{MEHS} the subjects' speech signals were used), spectrum density measurements must be made.

Studies completed so far have indicated that for talker sidetone at least, the rating method which correlates best with subjective effects of sidetone is one which takes into account the human sidetone signal as a masking threshold, i.e. sidetone masking rating (STMR).

3.2 *Listener sidetone*

3.2.1 *Definition of listener sidetone rating (LSTR)*

When the subscriber is listening, any room noise may reach the ERP through paths a) and c) of Figure 3/P.76. It is the high frequencies of local room noise which are most likely to mask the low-level consonants of a received signal. The STMR method described in § 3.1 has the effect of controlling L_{meST} more effectively at frequencies higher than 1000 Hz. Control of these frequencies is also important for room noise sidetone. This is because the low frequencies of a received signal at the earphone will be masked by low frequency room noise (leaking past the earcap) in much the same way as the talker's speech signal heard via the telephone sidetone path (L_{meST}) is masked by that heard via the human sidetone path (L_{MEHS}).

Studies have shown that if the room noise sidetone path (L_{RNST}) is determined as described in Recommendation P.64, and used in the STMR rating method, the resulting ratings correlate well with the subjective effects of room noise heard over the telephone sidetone path. The explanation of this is that the composite room noise signal arriving at the listener's ear and which performs a masking function on the received speech signals is believed to have a characteristic very similar to that of L_{MEHS}.

Thus LSTR is defined as that attenuation that must be inserted into the IRS (Recommendation P.48) to give an equivalent loudness to L_{RNST} when similarly taking L_{MEHS} into account as a masking threshold (Recommendation P.79).

3.2.2 *Determination of LSTR*

To calculate LSTR it is necessary to determine the sensitivity S_{RNST} (where $S_{RNST} = -L_{RNST}$) using a method such as that described in Recommendation P.64, or in the *Handbook on Telephonometry*, Section 3, and making use of the calculation procedure given in Recommendation P.79.

S_{RNST}, room noise sidetone sensitivity, will, in general, not have the same value as S_{meST}, talker sidetone sensitivity, since the sensitivity of the handset microphone may not be the same for random incidence signals as for a point source close to the diaphragm (less that 5 cm). Usually room noise arrives at the microphone at lower levels than speech and this can result in different sensitivity values, particularly where carbon microphones are present.

The difference between S_{RNST} and S_{meST} for a given telephone will usually be constant for different line conditions provided that it is operating in a linear part of its characteristic, and/or the room noise level is constant. This difference is Δ_{Sm}, (or DELSm), and is explained further in Recommendations P.10 and P.64, § 9. The use of Δ_{Sm} can be convenient where values of S_{meST} are known, to determine S_{RNST} for the purpose of calculating LSTR. Thus:

$$S_{RNST} = S_{meST} + \Delta_{Sm}$$

Normally Δ_{Sm} is negative, thus telephones that have a more negative value for Δ_{Sm} will have a lower value of S_{RNST} and perform better in noisy room conditions from the point of view of sidetone.

For telephone sets with linear microphones, Δ_{Sm} can vary over several decibels, typical values ranging from -1.5 to -4 dB. For carbon microphones, measurement values have been reported as low as -15 dB at some frequencies, but typical average values probably lie in the region of -8 dB for a room noise of 60 dBA. For some sets with linear microphones, the gain is intentionally not constant over their input/output characteristics in order to improve performance in noisy conditions. (See also Recommendation G.111, Annex A on the subject of Δ_{Sm}).

Note – Supplement No. 11 provides information on some of the effects of sidetone on transmission performance quantified over a number of study periods.

ANNEX A

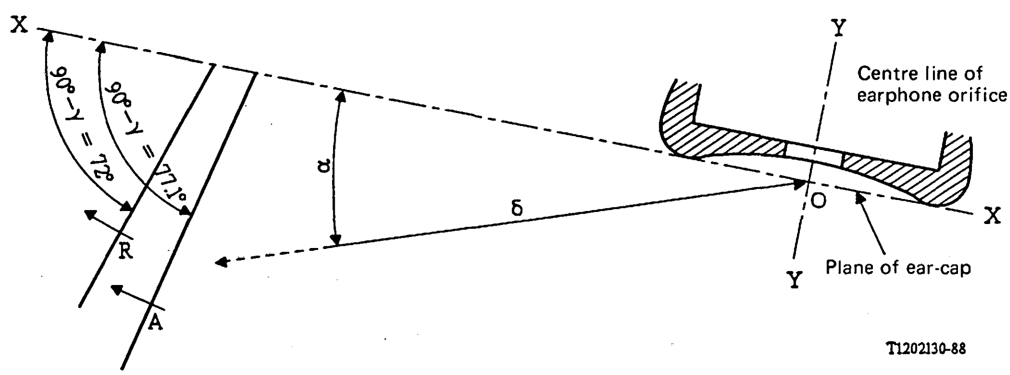
(to Recommendation P.76)

Definition of the speaking position for measuring loudness ratings of handset telephones

This annex describes the speaking position which should be used to measure the sensitivities of commercial telephone sets (by the method described in Recommendation P.64) for the determination of loudness ratings.

A.1 The definition of a speaking position falls into two parts: description of the relative positions of mouth opening and ear-canal opening on an *average* human head; and description of the angles that define the attitude in space of telephone handsets held to such a head. For any given telephone handset, these descriptions together describe the relative special disposition of the microphone opening and the talker's lips, and hence the direction in which speech sound waves arrive at the mouthpiece and the distance they have travelled from a *virtual point source.*

The relative positions of the centre of the lips and that of the ear canal can be described in terms of a distance δ and an angle α as shown in Figure A-1/P.76. Point R in that figure represents the centre of a guard ring located at the reference equivalent speaking position in accordance with Recommendation P.72, *Red Book.* Position A is that used to determine ratings by the articulation method defined in Recommendation P.45, *Orange Book.* Averages of lip positions of 4012 subjects in the People's Republic of China cluster round the point A (see Recommendation P.35).



Note 1 – Points R and A are located as follows:

A) $\delta = 136$ mm, $\alpha = 22°$, $\gamma = 12.9°$
R) $\delta = 140$ mm, $\alpha = 15.5°$, $\gamma = 18°$.

Note 2 – Solid lines through A and R show plane of lips.

FIGURE A-1/P.76

Location of lip position relative to opening of ear canal

A second angle is required to define the direction in which speech is emitted from the mouth into the mouthpiece of the microphone. In former Recommendations P.45 and P.72 reference is made to an angle β, but this does not lie in the plane of symmetry of the handset, so it is more convenient to use an angle γ, which describes the vertical projection of the direction of speech on this plane.

A.2 The position of the centre of the lips as defined by A in Figure A-1/P.76 is used also to define the new speaking position, but two additional angles must also be defined, namely: the earphone rotational angle Φ and the handset rotational angle Θ. Earphone rotation is considered about an axis through the centre of the ear-cap (YY in Figure A-1/P.76); handset rotation is taken about a longitudinal axis of the handset (XX in Figure A-1/P.76); both angles are zero when the plane of symmetry of the handset is horizontal. Naturally, the earphone rotational angle is positive when the handle is pointed downwards away from the earphone and the handset rotational angle is positive in the sense that the upper part of the earphone is moved farther from the medial plane of the head.

The new speaking position is described by the following values for the distance and angles defined above:

$$\alpha = 22°, \quad \gamma = 12.9°, \quad \delta = 136 \text{ mm}, \quad \Phi = 39° \quad \text{and} \quad \Theta = 13°$$

The angle γ cannot be determined very precisely and is not convenient for use when setting up a handset for test in front of an artificial mouth. The semi-interaural distance ε may be used in its place, and for the new speaking position ε = 77.8 mm.

For any test jig, the manufacture tolerance should be within ± 0.5° for the angles defined above.

A.3 The foregoing description of the speaking position has shown the complexities of expressing the relative location of the ear reference point and the guard-ring centre, and the relative orientation of the earphone axis and the guard-ring axis. It is often more convenient, particularly in terms of constructing and setting up handset jigs, to express the position of the ear reference point[4)] and the direction of the earphone axis with respect to the lip-ring. This is easier since the axis of the guard-ring is horizontal as would be the axis of an associated artificial mouth.

A.4 Use has been made of a vector analysis method to determine the orthogonal coordinates of the handset ear-cap relative to the lip position when the handset is mounted in the LR guard ring position. It is necessary to define a set of cartesian axes with origin at the centre of the lips (or equivalent lip position of an artificial voice) as follows:

x-axis: horizontal axis of the mouth, with positive direction into the mouth;

y-axis: horizontal, perpendicular to the x-axis, with positive direction towards the side of the mouth on which the handset is held;

z-axis: vertical, with positive direction upwards.

The ear reference point is defined by the vector:

$$(86.5, 77.8, 70.5) \text{ mm.}$$

The handset is mounted so that the ear reference point lies at the intersection of the axis of the ear-cap with a plane in space on which the ear-cap can be considered to be resting. With some shapes of handset, this definition is not adequate; in such cases the position of the ear reference point relative to the handset should be clearly stated.

The orientation of the handset is defined by vectors normal to the plane of the ear-cap and the plane of symmetry of the handset:

Unit vector normal to plane of the ear-cap:

$$\pm (0.1441, -0.974, 0.1748)$$

Unit vector normal to plane of symmetry of the handset:

$$\pm (0.6519, -0.0394, -0.7572).$$

4) See Recommendation P.64 for definition of ear reference point.

When using an artificial voice, the equivalent lip position must be used as the datum; this is not normally the same as the plane of the orifice of the artificial mouth.

Alternatively, it can be convenient to define the speaking position in terms of axes with the origin at the ear reference point. These are defined as follows:

x-axis: axis of ear-cap with positive direction away from earphone;

y-axis: line of intersection of the plane of symmetry of the handset with the ear-cap plane, with positive direction towards the microphone;

z-axis: normal to the plane of symmetry of the handset with positive direction obliquely upwards.

The lip-ring centre is defined by the vector:

$$(50.95,\ 126.10,\ 0)\ \text{mm}.$$

The orientation of the lip-ring is defined by a unit vector along its axis:

$$\pm\,(0.1441,\ -0.7444,\ -0.6250)$$

and the orientation of the handset is defined by specifying the vertical by the unit vector:

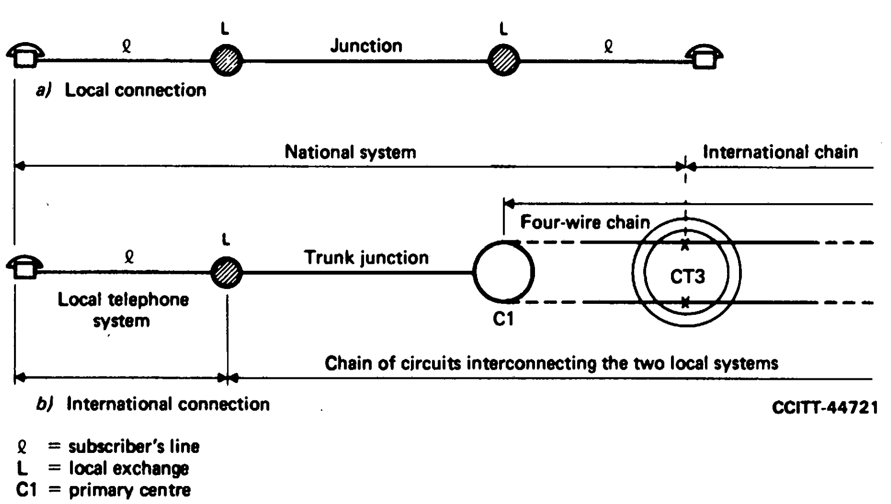
$$\pm\,(0.1748,\ -0.6293,\ +0.7572).$$

Note – The speaking position defined above differs from the special guard-ring position in the values of $\Phi\ (= 37°)$ and $\Theta\ (= 19°)$. It has been found that altering the handset position from the special guard-ring position to the loudness rating guard-ring position described above affects sensitivity measurements to a negligible extent.

ANNEX B

(to Recommendation P.76)

Explanations of certain terminology



ℓ = subscriber's line
L = local exchange
C1 = primary centre

FIGURE B-1/P.76

The terminology of Figure B-1/P.76 applies to parts of a telephone connection according to Recommendations G.101 [3], G.111 [4], G.121 [5] and CCITT manuals.

Note – In the present Recommendation the word "junction" is used in a special sense to denote "chain of circuits interconnecting the two local systems" and the "junction attenuator" used in laboratory tests for determination of loudness ratings.

References

[1] CCITT – Question 19/XII, Contribution COM XII-No. 1, Study Period 1985-1988, Geneva, 1985.

[2] CCITT Manual *Transmission planning of switched telephone networks*, Chapter I, Annex 1, ITU, Geneva, 1976.

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

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

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

Recommendation P.78

SUBJECTIVE TESTING METHOD FOR DETERMINATION OF LOUDNESS RATINGS IN ACCORDANCE WITH RECOMMENDATION P.76

(amended at Malaga-Torremolinos, 1984, Melbourne, 1988)

Preface

This Recommendation describes a subjective testing method which has been found suitable for its purpose by use in the CCITT Laboratory. It can also be used in other laboratories. Provided that the Intermediate Reference System (IRS) used complies with the requirements of Recommendation P.48 and that other requirements given in Recommendation P.76 are adhered to, the loudness ratings obtained by using the method given in the present Recommendation can be used for forwarding the study of Question 19/XII [1] (Recommended values of loudness rating). The present Recommendation, together with Recommendations P.76 and P.48, provides a definition of loudness ratings which can be used for planning.

Summary

This Recommendation contains the essential particulars for defining the method for determining loudness ratings in accordance with Recommendation P.76 when use is made of subjects performing equal loudness balances. Details are included concerning the balancing method, choice of subjects, speech material, design of experiment, method of analysis and presentation of results.

Study is continuing under Question 8/XII on using a direct-balance method. A description of this method can be found in Supplement No. 17.

1 Introduction

To compare the calculation of loudness ratings method (Recommendation P.79) a defined method of subjectively determining loudness ratings is required. This Recommendation deals with all aspects of a test from selection of operators to the method of analysis and finally presentation of results.

2 General

In the subjective comparisons, the Fundamental Reference System (FRS) is used (although other reference systems are permissible) as the datum for comparing the following speech paths:

a) *Path 0* – The fundamental reference system always provides the speech path against which each of the others is balanced. NOSFER set at 25 dB is used.

b) *Path 1* – The send end of the test ("unknown") local telephone circuit connected through the test ("unknown") junction and an adjustable attenuator to the receive end of the test ("unknown") local telephone circuit. The adjustable attenuator must be inserted in such a manner that the impedance relationships between the three parts of the connection (send end, junction and receive end) are not disturbed.

c) *Path 2* – The send end of the intermediate reference system connected through an adjustable attenuator to the receive end of the intermediate reference system.

d) *Path 3* – The send end of the test ("unknown") local telephone circuit connected through an adjustable attenuator to the receive end of the IRS.

e) *Path 4* – The send end of the IRS connected through an adjustable attenuator to the receive end of the test ("unknown") local telephone system.

f) *Path 5* – The send end of the IRS connected through the test ("unknown") junction and an adjustable attenuator to the receive end of the IRS. The adjustable attenuator must be inserted in such a manner that the impedance relationships between the three parts of the connection (send end, junction and receive end) are not disturbed.

In these subjective comparisons, the junction of the fundamental reference system is fixed, i.e. the level of speech sounds received via the fundamental reference system is kept constant, the loudness balance being obtained by the so-called "margin" method, and the balance attenuator being that inserted in the telephone (or IRS) path being tested.

The speaking position used with both the IRS and the test telephone sets should be as defined in Annex A to Recommendation P.76.

Figure 1/P.78 shows the composition of the telephone paths to be compared. The balances should be conducted using the vocal level defined in Recommendation P.72.

The loudness ratings relative to the IRS as defined in Recommendation P.76 are:

$$\mathrm{OLR} = x_2 - x_1$$

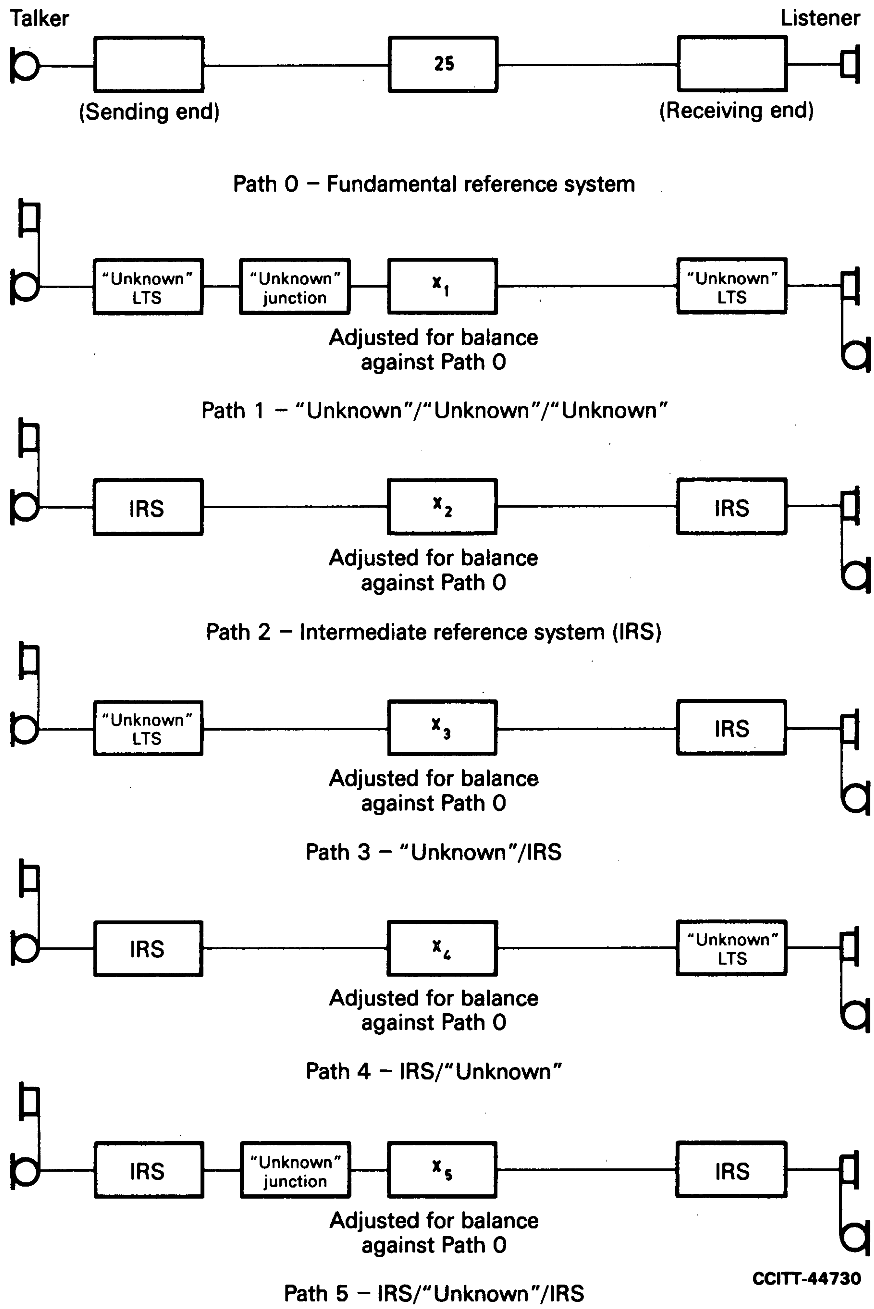
$$\mathrm{SLR} = x_2 - x_3$$

$$\mathrm{RLR} = x_2 - x_4$$

$$\mathrm{JLR} = x_2 - x_5$$

It is not necessary to include all the paths indicated above in every experiment. Paths 0 and 2 are essential but addition of only 3 and 4 is sufficient to determine sending and receiving loudness ratings of a local telephone circuit. Paths 0, 2 and 5 are required to determine a junction loudness rating. Path 1 is usually required only when it is derived to verify additivity of loudness ratings, namely that:

$$\mathrm{OLR} = \mathrm{SLR} + \mathrm{JLR} + \mathrm{RLR}$$



Note – Direct loudness balance of the "unknown" system against the IRS is also possible (see Supplement No. 17).

FIGURE 1/P.78

Arrangement of paths for subjective method of determination of loudness ratings

3 Experiment design

To have confidence in results requires the correct testing procedures to be followed, coupled with the correct experiment design. The procedure should be prepared such that no ambiguity can exist.

The following points must be considered in the design:

a) The experiment should be designed in such a way that all uncontrolled influences operate at random, e.g. slight day-to-day drift of subjects and/or measuring equipment;

b) If more balances are required than can be comfortably completed in one day, then the experiment must be designed such that equal numbers of each type of system are completed each day;

c) The operators who start a test should always be the same throughout the test [2];

d) A minimum of 12 operator-pair combinations is suggested with a maximum of 20. Twelve operator-pair combinations can be arrived at from two crews of 3 (see Table 1a/P.78) or one crew of 4 and 18 operator-pair combinations can be arrived at from one crew of 6 (see Table 1b/P.78) and 20 operator-pair combinations from one crew of 5 (see Table 2a/P.78).

Note – One crew of 6 giving 30 operator-pair combinations (see Table 2b/P.78) produces a larger test for only slightly more precision than the previously mentioned crew sizes;

TABLE 1a/P.78

Twelve operator-pair combinations from two crews of three, known as 3/6 operator method

		Operator (listener)					
		A	B	C	D	E	F
Operator (talker)	A		X	X			
	B	X		X			
	C	X	X				
	D					X	X
	E				X		X
	F				X	X	

TABLE 1b/P.78

Eighteen operator-pair combinations from one crew of six, known as 3/6′ operator method

		Operator (listener)					
		A	B	C	D	E	F
Operator (talker)	A				X	X	X
	B				X	X	X
	C				X	X	X
	D	X	X	X			
	E	X	X	X			
	F	X	X	X			

TABLE 2a/P.78

Twenty operator-pair combinations from one crew of five, known as 5/5 operator method

		Operator (listener)				
		A	B	C	D	E
Operator (talker)	A		X	X	X	X
	B	X		X	X	X
	C	X	X		X	X
	D	X	X	X		X
	E	X	X	X	X	

TABLE 2b/P.78

Thirty operator-pair combinations from one crew of six, known as 6/6 operator method

		Operator (listener)					
		A	B	C	D	E	F
Operator (talker)	A		X	X	X	X	X
	B	X		X	X	X	X
	C	X	X		X	X	X
	D	X	X	X		X	X
	E	X	X	X	X		X
	F	X	X	X	X	X	

e) When using two crews of 3, one can use both crews interleaved but it is generally more practical to separate the crews and use test crew 1 before crew 2. Members should not be used in both crews as it causes a bias and complicates the analysis;

f) All operator-pair combinations should be tested in rotation, where practical, such that each operator takes a turn as talker, then listener and then has a break;

g) The design of the experiment should eliminate any effect that could be attributed to the order of presentation. That is to say that all systems should be in a randomized order. To illustrate this point two examples are as follows:

Example 1

If one type of loudness rating is required, with a given combination of telephone set and circuit condition, then the experiment design must allow for any effect associated with order of presentation for each operator-pair combination. An example is shown in Table 3/P.78.

Note – However, if a laboratory has found with sufficient evidence that this method of design is not necessary, then a simplified design may be used.

TABLE 3/P.78

Example to illustrate the elimination of order of presentation effect for one type of loudness rating

Operator-pairs	Talker Listener	A B	B C	C A
Circuits	α α' β β'	3 2 1 4	1 3 4 2	2 4 3 1

Where

α = path 0 presented before path trajet 2

α' = path 2 presented before path trajet 0

β = path 0 presented before path trajet 3

β' = path 3 presented before path trajet 0

Note – When it is proven that there is no difference for a given test crew and set of test conditions, the distinction between the order of path presentation can be eliminated.

Example 2

Now, if more than one type of loudness rating is made or more than one telephone set is used, then there need only be one balance of path 2 against path 0 and vice-versa per operator-pair combination for any experiment, but this must be randomized within the experiment. An example is shown in Table 4/P.78.

TABLE 4/P.78

Example to illustrate the elimination of order of presentation effect for two type of loudness rating

Operator-pairs	Talker Listener	A B	B C	C A
Circuits	α	3	1	2
	α'	5	4	6
	β_1	1	2	5
	β'_1	6	5	3
	β_2	2	6	4
	β'_2	4	3	1

β_1, β'_1 = have, for example, 0 km of subscriber's cable

β_2, β'_2 = have, for example, 6 km of subscriber's cable.

Some experiment designs can be found in Annex A.

4 Selection of crew members and speech material

Requirements for the selection of crew members including audiometric testing of subjects, as well as the speech material used by the crew for subjective tests, can be found in Annex B.

5 Calibration of the IRS

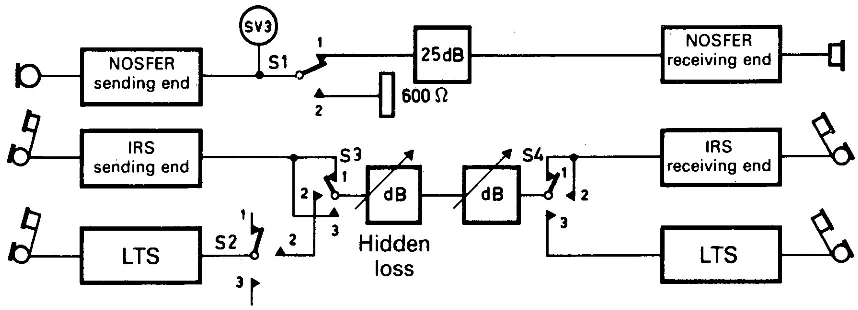
It is most important that the calibration of the IRS is made before every test so that any small change in SLR and RLR can either be compensated for in the results or the sensitivity can be changed before the test. It is good experimental practice to check the sensitivity of the IRS after each experiment. The specification of the IRS is found in Recommendation P.48 and the description of the calibration procedure is found in Recommendation P.64. The results of the calibration are used to determine the corrections to the subjective balance results (see § 9).

6 Circuit arrangements

Figure 2a)/P.78 shows a typical circuit layout for the measurement of SLR and RLR. Figures 2b)/P.78 and 2c)/P.78 show layouts for the measurement of JLR and OLR respectively. There is no reason if the experimenter so wished, why all four types of loudness rating cannot be tested in the same experiment. This, however, would require extremely intricate switching arrangements.

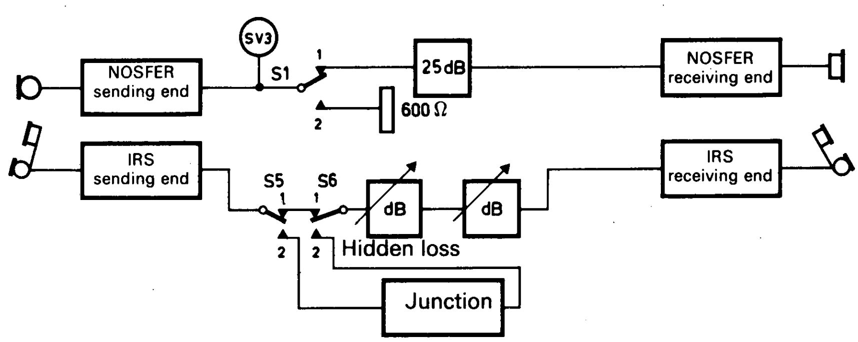
In Figures 2a)/P.78, 2b)/P.78 and 2c)/P.78 the 600 ohm on the second position of switch S1 allows the correct speech level to be set when Path 0 is presented after Path 1/2/3/4/5 (see Figure 1/P.78). This switch should be of the nonlocking type and should be returned to the normal position as soon as the talker has attained the correct speech level.

In order to reduce the effect of sidetone on the talker's vocal level during sending and overall determinations, the acoustic sidetone path of handset telephones should be disabled. This can be accomplished by placing the earphone in another identical handset and the electrical connections made to the correct terminals on the telephone transmission circuit. The earphone can then be sealed to an IEC/CCITT artificial ear to give the correct acoustic loading. A simpler method, used by the Australian Post Office, is to seal the earphone by means of heavy tape. Although this might not have the correct acoustic loading, in practice it has been found to have a negligible effect.



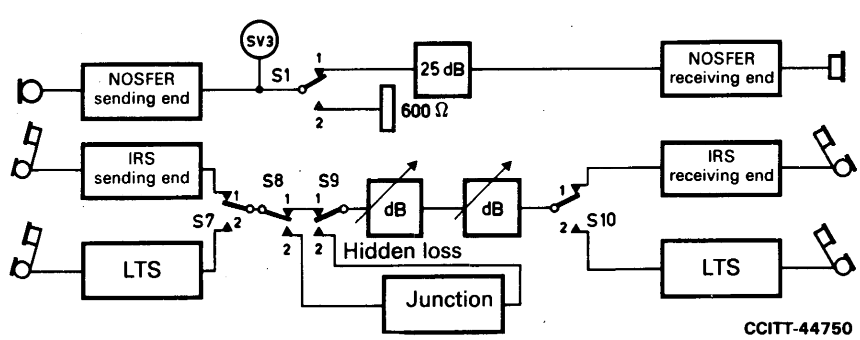
Note – S1 is a nonlocking switch. S2, S3 and S4 are all ganged.

a) Switching diagram for the measurement of SLR and RLR



Note – S1 is a nonlocking switch. S5 and S6 are ganged.

b) Switching diagram for the measurement of JLR



Note – S1 is a nonlocking switch. S7, S8, S9 and S10 are all ganged.

c) Switching diagram for the measurement of OLR

FIGURE 2/P.78

If the microphone is of the carbon-granule type, then before each balance the conditioning procedure according to Recommendation P.75 should be used.

In Figures 1/P.78 and 2/P.78 the fundamental reference system, NOSFER, has been shown but other types such as SETED and METRE-AIR-PATH could be used.

7 Recording of information

It is essential that as much information of any test should be recorded, in such a way that at any time in the future, the information can be retrieved.

7.1 *Details of the test*

Each test should always include the following information:

a) test No. – this should be unique so that one test cannot be confused with another;

b) date;

c) title – a brief description of the test;

d) circuit conditions – describe each individual path;

e) diagram to show switching arrangement;

f) crew members – name each operator and assign a code, as for example in Table 5/P.78. Then each operator-pair combination can be denoted by a code e.g. A-B.

TABLE 5/P.78

Crew members	
Code	Operator
A B C D E F	

7.2 *Individual balances*

These should always include the "hidden loss" attenuation, the "balance" attenuation and finally the result of the comparison, e.g.

$$R = H + B$$

where

R is the result

H is the hidden loss

B is the balance

8 Analysis

For any experiment most information can be obtained from an analysis of variance. However, sufficient useful information can be derived using the mean, standard deviation. The method of calculation of these parameters can be found in Annex C.

9 Presentation of results

The results of the test should be presented such that the important information can be displayed on one form. An example of such a form is shown in Table 6/P.78.

Note – In Tables 6/P.78 to 8/P.78 corrected mean = mean + correction.

Worked examples of the use of the form shown in Table 6/P.78 are shown in Tables 7/P.78 and 8/P.78. The form has been modified to allow SLR and RLR determinations to be made on a local telephone system including two line lengths. Table 7/P.78 shows the SLR results and Table 8/P.78 the RLR results.

TABLE 6/P.78
Presentation of results

Frequency Hz	IRS sending sensitivity (dBV/Pa)	IRS receiving sensitivity[a)] (dBPa/V)	Operator-pair	x_0 (dB)	x_2 (dB)	x_2' (dB)	x_3 (dB)	x_3' (dB)	x_2 (dB)	x_2' (dB)	x_4 (dB)	x_4' (dB)	SLR (dB)	SLR′ (dB)	RLR (dB)	RLR′ (dB)	$\frac{\text{SLR} + \text{SLR}'}{2}$ (dB)	$\frac{\text{RLR} + \text{RLR}'}{2}$ (dB)
100																		
125																		
160																		
200																		
250																		
315																		
400																		
500																		
630																		
800																		
1000																		
1250																		
1600																		
2000																		
2500																		
3150																		
4000																		
5000																		
6300																		
8000																		
Calculated LR of IRS			Mean: dB															
			Std. dev.: dB															
			95% confidence limits: dB															
				Corrected mean: dB														

a) Artificial ear conforming to Recommendation P.51.

TABLE 7/P.78

Example to illustrate the use of the form shown in Table 6/P.78 for the determination of SLR

Frequency Hz	IRS sending sensitivity (dBV/Pa)	IRS receiving sensitivity [a] (dBPa/V)	Operator-pair	x_0 (dB)	x_2 (dB)	x_2' (dB)	x_3 (0) (dB)	x_3' (0) (dB)	x_2 (dB)	x_2' (dB)	x_3 (L) (dB)	x_3' (L) (dB)	SLR (0) (dB)	SLR′ (0) (dB)	SLR (L) (dB)	SLR′ (L) (dB)	$\frac{SLR + SLR'}{2}$ (0) (dB)	$\frac{SLR + SLR'}{2}$ (L) (dB)
100			A-C	25	14	15	13	14			12	10	1	1	2	5	1.0	3.5
125			D-A	25	13	13	8	10			10	11	5	3	3	2	4.0	2.5
160			C-D	25	10	11	7	11			10	11	3	0	0	0	1.5	0.0
200	−19.7		D-C	25	12	14	11	10			10	11	1	4	2	3	2.5	2.5
250	−15.3		C-A	25	17	17	17	13			12	14	0	4	5	3	2.0	4.0
315	−12.2		A-D	25	10	12	8	10			10	8	2	2	0	4	2.0	2.0
400	− 9.6		F-E	25	11	11	7	7			5	4	4	4	6	7	4.0	6.5
500	− 8.0		B-F	25	10	11	6	8			5	7	4	3	5	4	3.5	4.5
630	− 6.7		E-B	25	13	12	8	13			8	9	5	−1	5	3	2.0	4.0
800	− 5.9		E-F	25	13	13	12	11			12	8	1	2	1	5	1.5	3.0
1000	− 5.6		F-B	25	12	13	9	5			5	6	3	8	7	7	5.5	7.0
1250	− 4.2		B-E	25	12	13	9	9			9	10	3	4	3	3	3.5	3.0
1600	− 1.2																	
2000	0																	
2500	+ 1.0																	
3150	+ 0.3																	
4000	−36.5																	
5000																		
6300																		
8000																		
Calculated LR of IRS	1.09		Mean: dB	25	12.25	12.92	9.58	10.08			9.00	9.08	2.67	2.83	3.25	3.83	2.75	3.54
			Std. dev.: dB	0	1.92	1.71	3.01	2.50			2.58	2.56	1.60	2.23	2.24	1.91	1.28	1.82
			95% confidence limits: dB	0	1.22	1.08	1.91	1.59			1.64	1.63	1.02	1.42	1.42	1.21	0.81	1.16
				Corrected mean: dB									3.76	3.92	4.34	4.92	3.84	4.63

[a] Artificial ear conforming to Recommendation P.51.

TABLE 8/P.78

Example to illustrate the use of the form shown in Table 6/P.78 for the determination of RLR

Frequency (Hz)	IRS sending sensitivity (dBV/Pa)	IRS receiving sensitivity [a] (dBPa/V)	Operator-pair	x_0 (dB)	x_2 (dB)	x_2' (dB)	x_4 (0) (dB)	x_4' (0) (dB)	x_2 (dB)	x_2' (dB)	x_4 (L) (dB)	x_4' (L) (dB)	RLR (0) (dB)	RLR′ (0) (dB)	RLR (L) (dB)	RLR′ (L) (dB)	$\frac{\text{RLR} + \text{RLR}'}{2}$ (0) (dB)	$\frac{\text{RLR} + \text{RLR}'}{2}$ (L) (dB)
100			C-B	25	10	11	20	20			15	13	−10	− 9	−5	−2	− 9.5	−3.5
125			B-E	25	15	9	19	21			13	13	− 4	−12	2	−4	− 8.0	−1.0
160			B-C	25	14	17	23	23			17	14	− 9	− 6	−3	3	− 7.5	0.0
200		− 3.8	E-B	25	11	10	19	19			13	15	− 8	− 9	−2	−5	− 8.5	−3.5
250		2.0	C-E	25	8	11	16	18			14	15	− 8	− 7	−6	−4	− 7.5	−5.0
315		6.6	E-C	25	13	13	18	18			13	16	− 5	− 5	0	−3	− 5.0	−1.5
400		9.8	D-F	25	8	9	13	13			12	9	− 5	− 4	−4	0	− 4.5	−2.0
500		11.2	F-A	25	14	14	22	21			17	16	− 8	− 7	−3	−2	− 7.5	−2.5
630		12.1	D-A	25	12	10	18	18			13	13	− 6	− 8	−1	−3	− 7.0	−2.0
800		12.8	A-D	25	12	8	21	19			12	11	− 9	−11	0	−3	−10.0	−1.5
1000		13.4	A-F	25	10	9	15	18			9	9	− 5	− 9	1	0	− 7.0	0.5
1250		13.8	F-D	25	11	9	19	16			10	10	− 8	− 7	1	−1	− 7.5	0.0
1600		14.0																
2000		13.2																
2500		11.0																
3150		10.4																
4000		−15.8																
5000																		
6300																		
8000																		
Calculated LR of IRS		− 0.16	Mean: dB	25	11.50	10.83	18.58	18.67			13.17	12.83	− 7.08	− 7.83	−1.67	−2.00	− 7.46	−1.83
			Std. dev.: dB	0	2.18	2.51	2.75	2.46			2.30	2.44	1.89	2.23	2.46	2.12	1.51	1.56
			95% confidence limits: dB	0	1.38	1.59	1.75	1.56			1.46	1.55	1.20	1.42	1.56	1.35	0.96	0.99
				Corrected mean: dB									− 7.24	− 7.99	−1.83	−2.16	− 7.62	−1.99

[a] Artificial ear conforming to Recommendation P.51.

ANNEX A

(to Recommendation P.78)

Examples of experiment designs

Tables A-2/P.78, A-3/P.78 and A-4/P.78, give typical designs for different crew sizes.

As an example, using Table A-2/P.78, the order of balances is as given in Table A-1/P.78.

TABLE A-1/P.78

Balance No.	Operator-pair	Circuit
1	BA	β_1
2	CB	α
3	DC	β_2
|		|
|		|
|		|
13	BA	β'_1
14	CB	β_1
15	DC	β'_2
|		|
|		|
|		|
25	BA	β_2
26	CB	β'_2
27	DC	α
|		|
|		|
|		|
71	AC	β_1
72	DA	α'

The operator-pairs in rotation do all balances in numerical order starting with "1" and finishing with "6".

Similar tables can be drawn up for a test requiring only one type of loudness rating where only 4 circuits are required e.g. α, α', β and β' for a SLR test, where numbers 1, 2, 3 and 4 would be assigned respectively in the experiment design.

For a test involving more circuits the same principles can be followed assigning as many numbers as there are circuits.

It may be necessary to improve the validity of results and a replication of the same experiment design using the same operator-pairs can be made.

TABLE A-2/P.78

Design for one crew of 4 or two crews of 3

One crew of 4	Talker	B	C	D	A	C	B	A	B	C	D	A	D
	Listener	A	B	C	D	A	D	B	C	D	B	C	A
Operator-pairs													
Two crews of 3	Talker	B	C	A	C	B	A	E	F	D	F	E	D
	Listener	A	B	C	A	C	B	D	E	F	D	F	E
Circuits	α	4	1	3	2	6	5	3	6	1	5	4	2
	α'	6	5	4	3	2	1	2	4	5	3	1	6
	β_1	1	2	5	6	3	4	5	3	2	1	6	4
	β'_1	2	4	6	5	1	3	4	2	3	6	5	1
	β_2	3	6	1	4	5	2	6	1	4	2	3	5
	β'_2	5	3	2	1	4	6	1	5	6	4	2	3

TABLE A-3/P.78

Design for one crew of 6

Opeertor-pairsateurs	Talker	D	E	F	E	F	D	F	D	E	A	B	C	A	B	C	A	B	C
	Listener	A	B	C	A	B	C	A	B	C	D	E	F	E	F	D	F	D	E
Circuits	α	4	1	3	2	6	5	3	6	1	5	4	2	1	2	6	3	5	4
	α'	6	5	4	3	2	1	2	4	5	3	1	6	5	4	1	6	2	3
	β_1	1	2	5	6	3	4	5	3	2	1	6	4	4	6	2	1	3	5
	β'_1	2	4	6	5	1	3	4	2	3	6	5	1	3	1	4	5	6	2
	β_2	3	6	1	4	5	2	6	1	4	2	3	5	6	5	3	2	4	1
	β'_2	5	3	2	1	4	6	1	5	6	4	2	3	2	3	5	4	1	6

TABLE A-4/P.78

Design for one crew of 5

Operator-pairs	Talker	B	C	D	E	A	C	E	B	D	A	D	B	E	C	A	E	D	C	B	A
	Listener	A	B	C	D	E	A	C	E	B	D	A	D	B	E	C	A	E	D	C	B
Circuits	α	4	1	3	2	6	5	3	6	1	5	4	2	1	2	6	3	5	4	1	6
	α'	6	5	4	3	2	1	2	4	5	3	1	6	5	4	1	6	2	3	2	5
	β_1	1	2	5	6	3	4	5	3	2	1	6	4	4	6	2	1	3	5	3	4
	β'_1	2	4	6	5	1	3	4	2	3	6	5	1	3	1	4	5	6	2	4	3
	β_2	3	6	1	4	5	2	6	1	4	2	3	5	6	5	3	2	4	1	5	2
	β'_2	5	3	2	1	4	6	1	5	6	4	2	3	2	3	5	4	1	6	6	1

ANNEX B

(to Recommendation P.78)

Selection of crew members, audiometric testing of subjects and speech material

B.1 *Crew members*

The crew should, wherever possible contain an equal number of both men and women.

The following points are a guide for selection:

a) Good hearing – no operator should exceed a hearing loss of a 15 dB at all frequencies up to and including 4 kHz and no more than 25 dB at 8 kHz. This is shown in Figure B-1/P.78. If it is intended that contra-lateral balances are required and this necessitates the use of both ears, then the maximum difference between ears should be ± 10 dB at all frequencies. An example of an audiometric testing procedure of subjects is presented below in § B.2;

b) Clear speech – each operator should be free from obvious speech impediments;

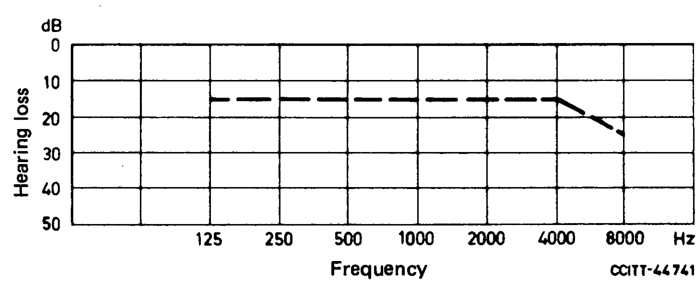
c) The operator should be able to work harmoniously with other people;

d) The operator should be able to make simple arithmetical calculations;

e) The operator should be able to talk at a constant level, with the aid of a meter, after sufficient training;

f) The operator must not suffer from claustrophobia as each operator must, during the test, spend a certain amount of short-term solitary confinement;

g) Regular checks should be made to determine the performance of each operator as both a talker and as a listener to disclose any unusual changes. A full description can be found in Reference [3].



Note – Nominal hearing is at 0 dB.

FIGURE B-1/P.78

Mask of maximum loss of hearing of subjects

B.2 *Audiometric testing of subjects – simple screening procedure [4]*

B.2.1 Visual examination of ears for wax, ask if subject has a cold, sinusitus or any other abnormality.

B.2.2 *Frequencies of test*

125, 250, 500, 1000, 2000, 3000, 4000, 6000, 8000 Hz.

B.2.3 *Example of presentation*

1000, 2000, 3000, 4000, 6000, 8000, 125, 250, 500, 1000 Hz.

Note – It is common for the second reading at 1000 Hz to be lower than the first.

Follow the above sequence for one ear, then repeat for the other ear.

B.2.4 *Example of finding threshold:*

Start above estimated threshold (say 20 dB hearing loss), approach in 10 dB steps until inaudible (no response). Return to last audible level and descend in 5 dB steps. Then approach this threshold from below in 5 dB steps. Signal duration 1 to 2 seconds.

Threshold is that value at which two equal responses are obtained from four successive stimuli.

B.2.5 *Room noise* [5]

Using supra-aural type headsets the maximum permissible levels in the test room are given in Table B-1/P.78.

If circum-aural type headsets are used then it is normally permissible to allow higher levels of noise.

TABLE B-1/P.78

Octave band	Sound pressure level (dB)
125	22.0
250	16.0
500	18.0
1000	26.0
2000	36.0
3000	39.5
4000	38.5
6000	40.0
8000	34.5

B.3 *Speech material*

The test phrase or phrases can be either a "nonsense" or "meaningful" phrase. Examples are:

a) Joe took father's shoe bench out,

b) Paris – Bordeaux – Le Mans – Saint-Leu – Léon – Loudun.

Due consideration should be given to the following points:

i) The ability of each operator to pronounce the chosen test phrase or phrases fluently and at a steady speech level. The sound structure of the native languages of the operators has therefore a bearing on the choice of test phrase or phrases;

ii) The phrase or phrases should be chosen so that the agreed measurement method to control the speech level (i.e. deflection of meter) can give a consistent and readily appreciated indication of vocal level.

ANNEX C

(to Recommendation P.78)

Simplified statistical analysis

C.1 *Mean*

The mean is obtained by using the following formula:

$$\bar{x} = \frac{\Sigma x}{n}$$

C.2 *Standard deviation*

It cannot be assumed that the operators are a sample drawn at random from a population and that the operator-pair combinations are independent of each other. Under these circumstances the standard deviation must be of the sample and not an estimate of a population.

The formula for the standard deviation is:

$$\sigma = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n}}$$

C.3 A more detailed statistical analysis is possible to calculate confidence intervals as explained in § 1.3.4 of the *Handbook on Telephonometry* [6]. The confidence interval is governed by the dispersion between the crew members, the number of crew members and the arrangement of the experimental design. Typical values in a well-conducted test are ± 5 dB for the arrangements shown in Table 1a/P.78, ± 4 dB for Table 1b/P.78, ± 3 dB for Table 2a/P.78 and ± 2 dB for Table 2b/P.78.

References

[1] CCITT – Question 19/XII, Contribution COM XII-No. 1, Study Period 1985-1988, Geneva, 1985.

[2] *The design and analysis of loudness efficacy measurements,* Red Book, Vol. V, Annex 7, ITU, Geneva, 1962.

[3] *Extract from a study of the differences between results for individual crew members in loudness balance tests,* Red Book, Vol. V, Annex 6, p. 214, ITU, Geneva, 1962.

[4] BURNS (W.): Noise and man, *Murray,* pp. 70-80, 1968

[5] *Ibid.,* pp. 298-300.

[6] CCITT – *Handbook on Telephonometry,* ITU, Geneva, 1987.

CALCULATION OF LOUDNESS RATINGS

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

Preface

The method given in this Recommendation is provisional for the reason stated in detail below, that its applicability to local telephone systems containing carbon microphones has not been confirmed beyond doubt. Nevertheless, Administrations which are studying Question 19/XII [1] (recommended values of loudness ratings) may use this method for studies relating to new types of telephone sets which do not contain carbon microphones [1].

Administrations are also encouraged to use the method in studying Question 7/XII [2] for expressing loudness loss on a common scale in quality evaluation experiments.

The Recommendation describes a calculation method which gives results in good agreement with those from subjective tests by the CCITT Laboratory [2] (see Recommendation P.78) using local telephone systems having noncarbon microphones. For such local telephone systems, the methods given in Recommendation P.64 should be used to determine the values of sending and receiving sensitivities.

When local telephone systems containing carbon microphones are to be considered, the results obtained so far from tests in the CCITT Laboratory suggest that the method given in the present Recommendation can still be used provided a suitable method is used to obtain the sending sensitivities. Various measuring methods are being considered for this purpose and are listed in Annex B to Recommendation P.64. Extensive tests by the CCITT Laboratory using the "upper-envelope" method show that this method gives good results for some types of carbon microphone. The matter is being studied under Question 8/XII [3] (Measuring the efficiency of a microphone or a receiver).

The loudness ratings of analogue telephone sets are determined objectively by special measuring instruments conforming to Recommendations P.64 and P.65 as regards the physical implementation, and to the present Recommendations as regards the computational algorithm. However, the results should not be applied directly for transmission planning, before certain precautions have been observed regarding bandwidth and terminating impedances.

1 Introduction

Loudness ratings according to the principles described in Recommendation P.76 can be determined without recourse to subjective tests provided that all the following conditions are fulfilled:

a) a theoretical model is available having suitable structure;

b) the appropriate values of the essential parameters of the model are known;

c) the sending and receiving sensitivities of the intermediate reference systems are known;

d) the sending and receiving sensitivities of the "unknown" local telephone systems and the insertion loss of the intervening chain of circuits are known.

The methods of determining sending and receiving sensitivities using an artificial mouth and artificial ear are defined in Recommendation P.64. The characteristics of the intermediate reference system determined according to the same methods are given in Recommendation P.48. The receiving sensitivities obtained using the artificial ear now mentioned in Recommendation P.64 are not directly suitable for use in calculating loudness ratings but must be corrected to allow for differences between sound pressures in real ears under conditions of telephone conversations and those measured by the artificial ear. Information concerning this correction (L_E) is given in § 6.

---

[1] The method may also be used for determining receiving loudness ratings whether or not the telephone set contains a carbon microphone.

[2] The calculation method described in the Recommendation is based on weighting factors which have been determined for the 20 ISO-preferred frequencies. General applicability of the method would be improved if smoothed analytic expressions were also available for use with other sets of frequencies.

2 Definitions and symbols concerning sound pressures, sensitivities and transmission losses

Definitions and symbols used in the subsequent description of theoretical principles are listed below. Figure 1/P.79 illustrates these.

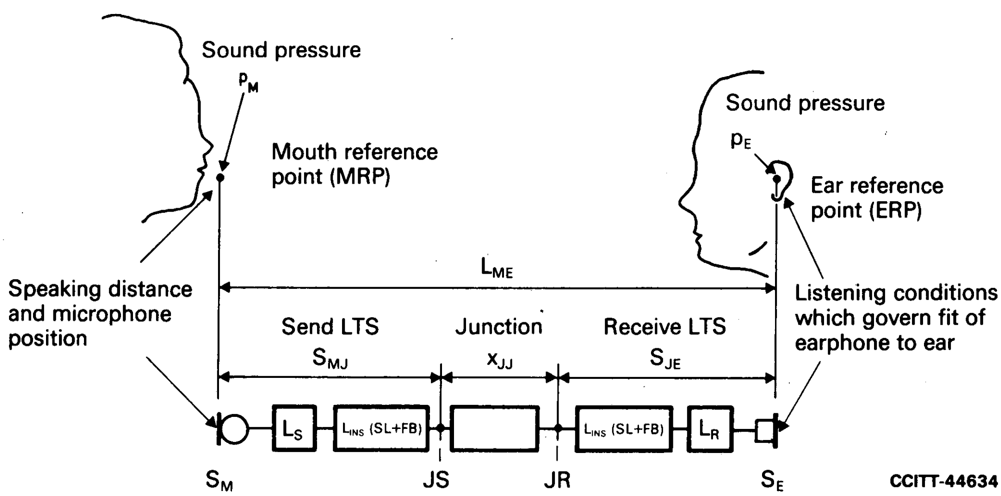


FIGURE 1/P.79
Factors effecting loudness of received speech

2.1 *Concerning talking*

These definitions and symbols characterize the situation where a subject is talking and they include his physical relationship to the telephone or reference connection.

MRP — Point defining the mouth reference point; MRP is at a defined location relative to the talker's lips. (See Recommendation P.64.)

p_M — Sound pressure at MRP[3] in absence of any obstruction.

B'_S — Spectrum density (long-term mean pressure)[4] of speech referred to a MRP in dB relative to 20 μPa in a bandwidth of 1 Hz.

VL — Vocal level, i.e. speech sound pressure (long-term rms while talker is active) level of talker at the MRP; usually referred to a reference vocal level as datum.

SP — Speaking position, i.e. the relative location of the microphone of the telephone or reference system and the lips of the talker.

---

[3] The reference level or datum must be specified, e.g. 1 Pa, 20 μPa, etc.

[4] In practice, measurements are made in terms of sound pressure, and that convention is retained for convenience of explanation. It is worth noting that sound pressure relative to 20 μPa in a bandwidth of 1 Hz is approximately equal to sound intensity relative to 1 pW/m² per Hz.

2.2 *Concerning listening*

These definitions and symbols characterize the situation where a subject is listening and they include his physical relationship to the telephone or reference connection:

ERP	Point defining the ear reference point (see Recommendation P.64).
p_E	Sound pressure at ERP.
β_0	Hearing threshold for pure tones referred to an ERP in dB relative to 20 μPa.
K	A number, related to Fletcher's critical frequency bands, required to convert hearing threshold for pure tones to that for continuous-spectrum sounds like speech.
$\beta_0 - K$	Hearing threshold for continuous-spectrum sounds referred to an ERP in dB relative to 20 μPa in a bandwidth of 1 Hz.
HL	Hearing loss, usually referred to "normal" hearing threshold.
LC	Listening conditions; the manner in which the earphone and its coupling to the ear is related to the ERP.

2.3 *Concerning telephone or reference connections*

These definitions and symbols serve to characterize the telephone or reference connections in objective terms:

L_{ME}	Air-to-air transmission loss, in dB, from a MRP to an ERP.
JS, JR	Electrical interfaces at the output of a sending local telephone system and the input to a receiving local telephone system.
LTC	Local telephone system.
S_{MJ}	Sending sensitivity of a local telephone system from the MRP to the electrical output (JS). *Note* – S_{MJ} relates to a median real mouth; for practical purposes, sensitivities measured according to Recommendation P.64 using the recommended artificial mouth may be used for handset telephones.
S_{JE}	Receiving sensitivity of a local telephone system from the electrical input (JR) to the ERP. *Note* – S_{JE} relates to a median real ear; sensitivities measured with the artificial ear referred to in Recommendation P.64 and according to the method described therein are denoted by the symbol S_{Je}. Such values must be corrected to give appropriate values for S_{JE} (see § 6).
x_{JJ}	Transmission loss between local telephone systems, i.e. between JS and JR in Figure 1/P.79. The circuits concerned in real telephone connections will consist of trunk junctions, trunk circuits, switching centres, etc. For assessment purposes this chain of lines is replaced by nonreactive attenuators and filters, etc. and referred to collectively by the word "junction".
S_{RMJ}, S_{RJE}, L_{RME}, etc.	Values of S_{MJ}, S_{JE}, L_{ME}, etc., applicable to a reference speech path, e.g. NOSFER or the IRS defined in Recommendation P.48.
S_{UMJ}, S_{UJE}, L_{UME}, etc.	Values of S_{MJ}, S_{JE}, L_{ME}, etc., applicable to an unknown speech path, e.g. a telephone connection.
x_{UR}, x_{RU}	Values of x applicable to combinations of "unknown" sending to reference receiving and reference sending to "unknown" receiving speech paths.
S_M	Sensitivity of a telephone microphone referred to a MRP.
S_E	Sensitivity of a telephone receiver referred to an ERP.
L_S	Electrical transmission loss from the terminals of a microphone to the line terminals of a telephone set.
L_R	Electrical transmission loss from the line terminals of a telephone set to the terminals of a receiver.
L_{INS} $(SL + FB)$	Transmission loss of the combination of subscriber's line and feeding bridge.

3 Structure of the theoretical model

3.1 *Definitions concerning loudness, its relationship to sensation level and loudness ratings*

These definitions and symbols relate to factors concerning loudness and loudness ratings of telephone speech paths:

Z	Sensation level, in dB, of the received speech signal at a given frequency; describes the portion of the received speech signal which is above threshold and is, therefore, effective in producing the sensation of loudness.
Z_{RO}	Value of Z when $L_{ME} = 0$ dB.
$Q(Z)$	Function of Z related to loudness; transforms sensation level expressed in terms of Z, to loudness numerics.
m	A parameter which can be used to define $Q(Z)$; represents the slope of $10 \log_{10} Q(Z)$ as function of Z.
S	A monotonic function of frequency such that equal increments of S are of equal importance to loudness, provided the associated values of Z are the same.
S'	The derivative of S with respect to frequency; $S' = \mathrm{d}S/\mathrm{d}f$. S' can be considered as a frequency weighting factor.
$\mathrm{d}S$	From the foregoing, $\mathrm{d}S = S'\,\mathrm{d}f$.
$\overline{Q(Z)}$	Weighted average of $Q(Z)$ which is related to the total loudness in a received speech signal.
λ	Loudness of the sound being considered.
OLR, SLR, RLR, JLR	Overall, sending and receiving and junction loudness ratings.

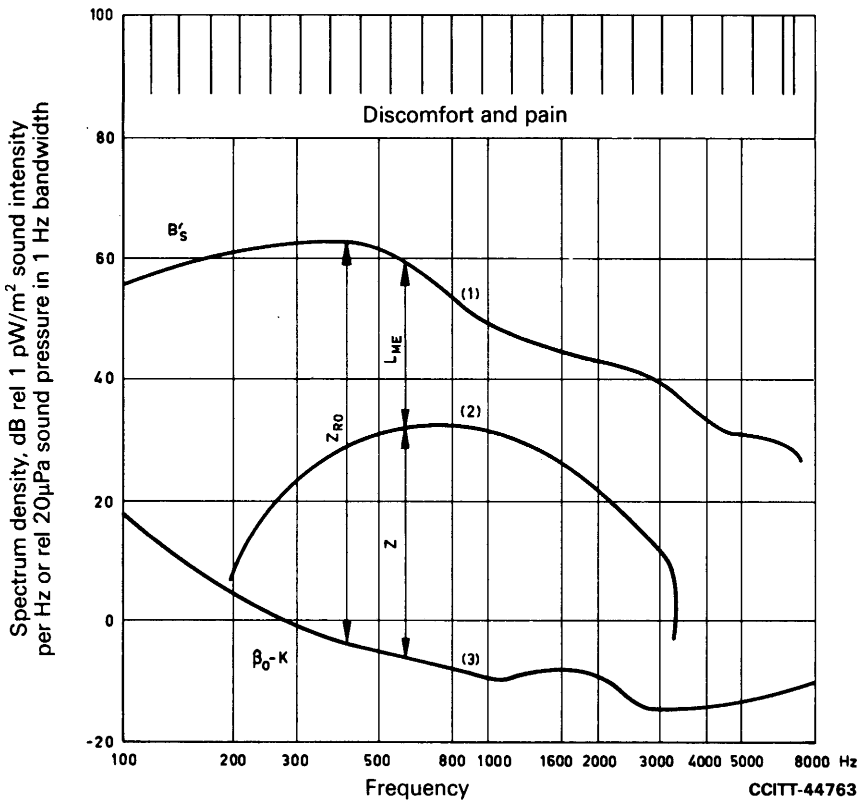
3.2 *Loudness model*

In considering speech transmission paths, it is necessary to define acoustical terminals of the paths. This can be done in terms of MRP and ERP. There are no unique definitions of such reference points, but those used here are defined in Recommendation P.64.

Curve 1 in Figure 2/P.79 shows the spectrum density B'_S of speech emitted at a certain vocal level and measured at the MRP in the absence of any obstruction in front of the mouth[5]. The measurement may be thought of as made with the aid of a very small measuring microphone. When the speech reaches the ear of the other participant in a telephone conversation, it will have been subjected to transmission loss and distortion in the telephone speech path and the spectrum density may then be as shown in Curve 2; the ERP to which Curve 2 is referred can, for explanation, be thought of as located at the opening of the ear canal, but might equally well be the tympanum, i.e. eardrum of the listener's ear. The studies at present in hand make use of an ear reference point located at the opening of the air canal (as referred to in Annex A to Recommendation P.64). The interval L_{ME} between curves 1 and 2 represents the "mouth-to-ear" transmission loss and is, in general, frequency-dependent.

The received spectrum represented by Curve 2 does not contribute uniformly to loudness, i.e. those portions of the spectrum lower in level than the listener's threshold of hearing contributes very little compared with those well above the threshold. Account is taken of this by defining a quantity termed "sensation level" (symbol Z) which is the interval between the received spectrum, Curve 2, and the threshold of audibility for continuous spectrum sounds ($\beta_0 - K$) shown in Curve 3. Loudness of the received speech sound thus depends upon Z, which is, in general, frequency-dependent.

[5] See Annex A to Recommendation P.64 for the definition of MRP.



Curve (1) Spectrum density of speech at mouth reference point.
Curve (2) Spectrum density of speech at ear reference point received over an approximately limiting telephone speech path.
Curve (3) Hearing threshold for continuous spectrum sounds.

FIGURE 2/P.79

Determination of sensation level Z, the portion of the received speech signal effective in producing the sensation of loudness

Studies have shown[6] that the loudness, λ, can be expressed approximately as a function of Z in the following manner:

$$\lambda = C \int_{f_1}^{f_2} Q(Z)\, S'\, \mathrm{d}f \tag{3-1}$$

where C is a constant, $Q(Z)$ is a "loudness growth function" which transforms Z so that equal increments of the transformed values represent equal increments in loudness, S' is a "frequency weighting function" which weights the transformed values of Z according to their positions along the frequency scale and f_1 and f_2 correspond to the lower and upper frequency limits for the band of interest.

[6] This model does not claim to represent accurately all the features that relate to perception of the loudness of speech; for example, the effects of interfrequency masking are ignored and it does not predict the increasing importance of the lower frequencies as the intensity of the sound is increased from the threshold. It is possible to construct models that represent more of the features fairly well, but no completely comprehensive model is known. Such models are unnecessarily complicated for calculating loudness ratings. The most important restriction with respect to this model is that it should be used to make comparisons at the constant listening level indicated in Recommendation P.76.

If desired the frequency scale can be transformed to a scale of S, equal increments of which have the same "importance" so far as loudness is concerned.

Thus:

$$S' = \frac{\mathrm{d}S}{\mathrm{d}f} \tag{3-2}$$

which gives

$$\lambda = C \int_{S_1}^{S_2} Q(Z)\,\mathrm{d}S \tag{3-3}$$

where S_1 and S_2 are points on the scale of S that correspond respectively to f_1 and f_2.

The basic elements of the loudness rating process are shown in the flow diagram of Figure 3/P.79. The flow diagram depicts a "reference" spectrum decreased by the loss of a telephone connection resulting in a received spectrum which together with the threshold of hearing produces Z, the values of which (as a function of frequency) are effective in producing the sensation of loudness. Thus:

$$Z = B'_S - L_{ME} - (\beta_0 - K) \tag{3-4}$$

and Z as a function of frequency is converted to loudness, λ, according to the equations explained above in which Z is transformed to loudness numerics which are then weighted by the frequency weighting function to produce $\overline{Q(Z)}$; a constant applied to $\overline{Q(Z)}$ produces λ, the loudness of the received speech expressed on some suitable scale.

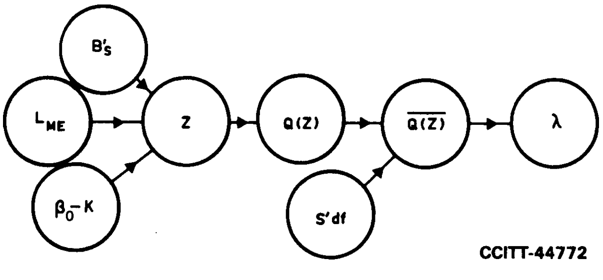


FIGURE 3/P.79

Simplified flow diagram showing how loudness, λ, is related to sensation level, Z

The flow diagram of Figure 3/P.79 represents only basic elements in the loudness rating process. These elements require further specification in order to render them unique. For example, B'_S depends on the particular speaker and his vocal level, the test phrase used, and the location of the talkers lips with respect to the telephone microphone defined by his individual method of usage and by the somewhat arbitrarily defined MRP. Similarly, the received spectrum level depends on the particular listener and his characteristics, e.g. fit between his ear and the telephone earphone when the handset is held in a prescribed manner, whether or not he has a hearing loss, and on the ERP.

Furthermore, transmission planning studies require subdivision of the connection loss, L_{ME}, into component parts, e.g. a sending component, a receiving component and an interconnecting component.

The function $Q(Z)$ can, in part, be specified in terms of a parameter m which is the slope of the logarithm of $Q(Z)$ when plotted against Z. m does, however, depend upon the listening level (or Z) in the general case but may be considered constant over a wide and useful range of Z.

Those additional factors considered at present to be of importance are included in the more detailed flow diagram of Figure 4/P.79 which is an expansion of Figure 3/P.79. The influence of these factors can be appreciated from the previous discussion and from review of the definitions given in § 3.1. Figure 3/P.79 supplements these definitions.

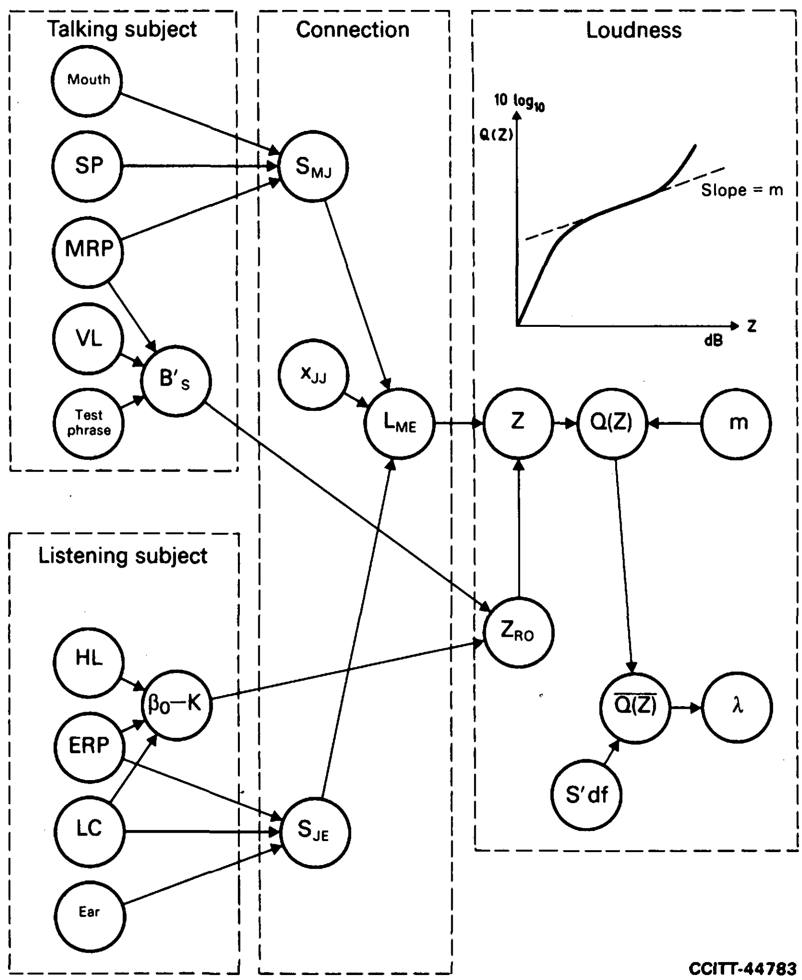


FIGURE 4/P.79
Flow diagram

4 Values of the parameters

4.1 *General*

To implement the model in the form described in § 3, it is, in principle, necessary to assign values to the following parameters:

B'_S as a function of frequency

$10 \log_{10} S'$ as a function of frequency

m which (partly) defines the loudness growth function $Q(Z)$

$\beta_0 - K$ as a function of frequency.

In fact, for the present purposes, it is convenient to group all these parameters together into a single frequency-dependent parameter which can be used with m for the purposes of calculating sending, receiving and junction loudness ratings and the loudness insertion loss of electrical elements such as channel filters in commercial telephone connections.

The theoretical derivation of this frequency-dependent parameter G, is explained below.

G, together with m, can be estimated directly from the results of subjective loudness balance tests conducted using sets of lowpass and highpass filters in a suitable reference system.

4.2 *Theoretical derivation of G*

Equation 3-1 can be written:

$$\lambda_U = C \int Q(Z_U)\, S'\, \mathrm{d}f \tag{4-1a}$$

and

$$\lambda_R = C \int Q(Z_R)\, S'\, \mathrm{d}f \tag{4-1b}$$

where λ_U and λ_R represent the loudness of speech received through the "unknown" and reference speech paths respectively and Z_U and Z_R are the corresponding values of sensation level (which are functions of frequency).

The calculation method to be described depends upon the assumption (largely verified for restricted ranges of listening level) that the function $Q(Z)$ can be put in the form:

$$Q(Z) = \text{constante} \cdot 10^{m(1/10)\,Z} \tag{4-2}$$

(The base 10 and the multiplier 1/10 are used merely to preserve the analogy to the decibel, in which unit Z is expressed.)

Let

$$Z_{RO} = B'_S - (\beta_0 - K) \tag{4-3}$$

and substitute in Equation 3-4 to obtain:

$$Z_U = Z_{RO} - L_{UME} \tag{4-4a}$$

$$Z_R = Z_{RO} - L_{RME} \tag{4-4b}$$

By substituting Equations (4-4a) and (4-4b) in Equations (4-1a) and (4-1b) and rearranging:

$$\lambda_U = C \int 10^{-m(1/10)\,L_{UME}} [10^{m(1/10)\,Z_{RO}}\, S']\, \mathrm{d}f \tag{4-5a}$$

$$\lambda_R = C \int 10^{-m(1/10)\,L_{RME}} [10^{m(1/10)\,Z_{RO}}\, S']\, \mathrm{d}f \tag{4-5b}$$

The loudness rating can be considered to be the Δx (independent of frequency) removed from the "unknown" speech path to render $\lambda_U = \lambda_R$.

Using the substitution:

$$G = [10^{m(1/10)\,Z_{RO}}\, S'] \tag{4-6}$$

and inserting $L_{UME} - \Delta x$ in Equation (4-5a) in place of L_{UME}, we obtain equality of the λ's.

Therefore

$$\int 10^{-m(1/10)(LUME-\Delta x)}\, G\mathrm{d}f = \int 10^{-m(1/10)LRME}\, G\,\mathrm{d}f \tag{4-7}$$

$$10^{-m(1/10)\Delta x} = \frac{\int 10^{-m(1/10)LUME}\, G\,\mathrm{d}f}{\int 10^{-m(1/10)LRME}\, G\,\mathrm{d}f} \tag{4-8}$$

and

$$\Delta x = -m^{-1}\, 10 \log_{10} \int 10^{-m(1/10)LUME}\, G\,\mathrm{d}f - \left\{-m^{-1}\, 10\log_{10} \int 10^{-m(1/10)LRME}\, G\,\mathrm{d}f\right\} \tag{4-9}$$

Without affecting the equality, G can be scaled by multiplying with a suitable constant to render $\int G\,\mathrm{d}f = 1$; G can then be treated as a weighting factor[7] and each term on the right-hand side takes the form:

$$\Phi^{-1}\left[\int \Phi(L)\, G\,\mathrm{d}f\right] = \overline{L}$$

Then for the loudness rating we have

$$\text{loudness rating} = \Delta x = \overline{L_{UME}} - \overline{L_{RME}} \tag{4-10}$$

The terms $\overline{L_{UME}}$ and $\overline{L_{RME}}$ can be considered as the "weighted average mouth to ear loss" of the "unknown" and reference speech paths respectively. In each of the foregoing equations, integration (and therefore averaging) is over the range between lower and upper frequency limits of interest.

For computation, the audible range of frequency is divided into a number (N) of continuous band; use is made here of the 20 ISO-preferred bands centred at frequencies spaced at approximately 1/3 octaves from 100 to 8000 Hz. Averaging the values of $\overline{L_{UME}}$ is then performed by summations of the form:

$$\overline{L_{UME}} = -m^{-1}\, 10\log_{10} \sum_{i}^{N} 10^{-m(1/10)LUME}\, G\,\Delta f \tag{4-11}$$

The acoustical transmission loss of a speech path is, in general, a function of frequency and can be defined as:

$$L_{UME} = 20\log_{10} \frac{p_M}{p_E} \tag{4-12}$$

where p_M and p_E are as defined in §§ 2.1 and 2.2.

It is necessary to know the values of L_{UME} at each frequency together with $G\Delta f$; naturally, L_{UME} depends on the telephone speech path under consideration but $G\Delta f$ and other information common to all speech paths is described below.

4.3 *Determination of values for G*

Values have been assigned to G by analysis of results of loudness balance tests by the CCITT Laboratory using a special speech path consisting of NOSFER, but with its sending frequency response made more level by equalization. Each of a set of special low- and high-pass filters was inserted in turn in the "junction" of this speech path.

[7] From Equations (4-3) and (4-6) it can be seen that G as a function of frequency depends upon the value of m and the frequency-dependent functions B'_S, β_0, K and S'.

Balances were made with each filter and with the "through" path; each was treated as the "unknown" while balancing for determining relative equivalents against NOSFER with its junction set at 25 dB. Balancing was done by the "margin" method, i.e. by changing the transmission loss in the "unknown". Values of Δx were calculated for each filter and corrected for the transmission loss in the pass-band. The cut-off frequencies were taken as those frequencies at which the transmission loss was 10 dB greater than the pass-band transmission loss.

By smoothing the results and interpolating at the appropriate edges of the 20 ISO-preferred frequency bands centred at the frequencies from 100-8000 Hz, it was possible, first, to estimate m; $m = 3/\Delta x$, if we take the value of Δx at the frequency where Δx was the same for low- and for high-pass filtering. Then, by use of Equation (4-8) and some interaction, it was possible to obtain a set of values for G which satisfied the experimental data. Note that L_{RME} in Equations (4-7) to (4-10) represents the mouth-to-ear transmission loss of the "through" path and L_{UME} represents that of the same path with the filter inserted.

The results are given in Table 1/P.79, the value determined for m being 0.175.

TABLE 1/P.79

Values of 10 $\log_{10}$ G and 10 $\log_{10}$ $G\,\Delta f$ determined by the CCITT Laboratory

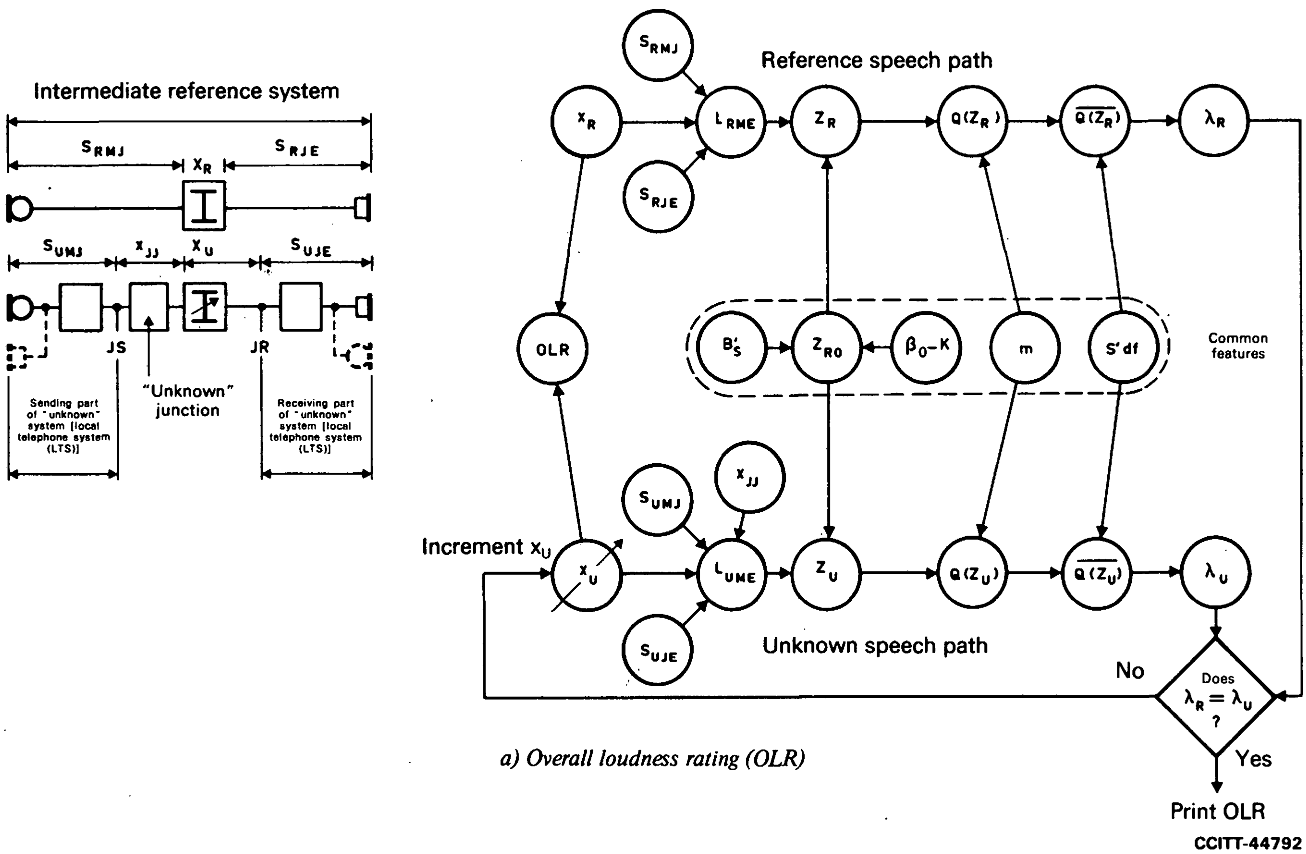
Midfrequency (Hz)	Δf (Hz)	10 $\log_{10}$ G (dB)	10 $\log_{10}$ $G\,\Delta f$ (dB)
100	22.4	−32.63	−19.12
125	29.6	−29.12	−14.41
160	37.5	−27.64	−11.90
200	44.7	−28.46	−11.96
250	57.0	−28.58	−11.02
315	74.3	−31.10	−12.39
400	92.2	−29.78	−10.14
500	114.0	−32.68	−12.12
630	149.0	−33.21	−11.48
800	184.0	−34.14	−11.49
1000	224.0	−35.33	−11.83
1250	296.0	−37.90	−13.19
1600	375.0	−38.41	−12.67
2000	447.0	−41.25	−14.75
2500	570.0	−41.71	−14.15
3150	743.0	−45.80	−17.09
4000	922.0	−43.50	−13.86
5000	1140.0	−47.13	−16.56
6300	1490.0	−48.27	−16.54
8000	1840.0	−46.47	−13.82

5 Calculation of loudness ratings

5.1 *Deviation of formulas and W weights*

The method described in Recommendation P.78 can be described in terms of the flow diagrams illustrated in Figure 5/P.79 which also embody the structure of the model used here (Figure 4/P.79). The diagrams placed on the left in parts a), b), c) and d) of Figure 5/P.79 are redrawn versions of the various paths given in Figure 1/P.78.

Figure 5/P.79 illustrates the procedure when values are known for all the parameters referred to in §§ 1, 2 and 3. In a) of Figure 5/P.79, the parameters shown grouped together are those used to form the composite parameter G described in § 4. Further grouping is possible as shown in b), c) and d) of Figure 5/P.79. It will also be seen that the whole of the path from x_R to λ_R is also common to all four flow diagrams. Use can be made of this feature to reduce the calculation procedure to a formula which is very easy to compute.



a) Overall loudness rating (OLR)

Note concerning a) of Figure 5/P.79

The "unknown" path consists of four sections as follows:

a) sending LTS, comprising telephone set, subscriber's line and feeding bridge, up to JS in Figure 1/P.79;

b) receiving LTS, comprising feeding bridge, subscriber's line and telephone set, from JR in Figure 1/P.79;

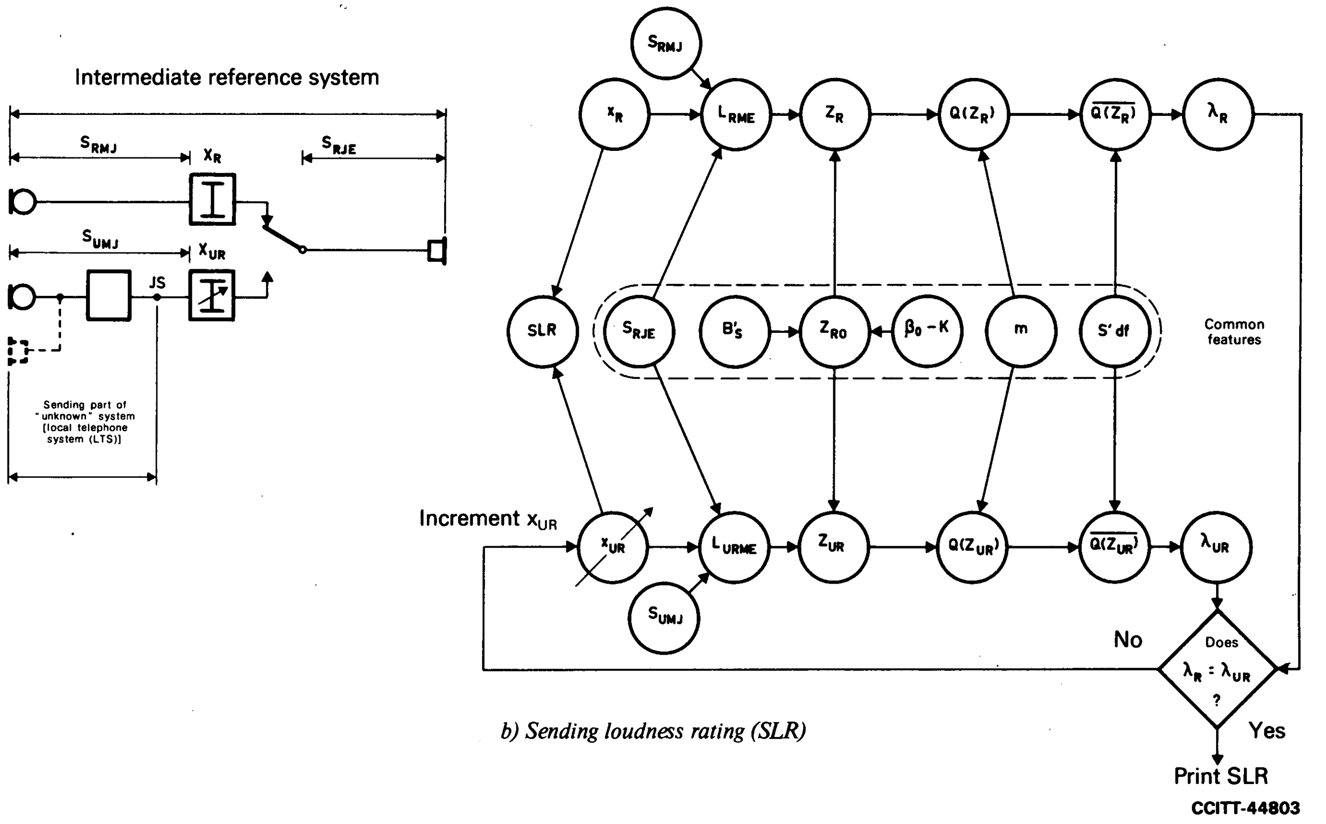
c) the combination of trunk junctions and trunk circuits present in the real connection between JS and JR;

d) additional, adjustable, transmission loss, x_U, introduced in such a manner that it will not disturb the overall frequency response of the complete connection, but will only increase the transmission loss equally at all frequencies.

If the section of the real connection between JS and JR has an image impedance of 600 ohms $\angle 0°$, there is no problem either in defining x_{JJ} or in introducing the additional loss, x_U. Where this is not so, the image attenuation of a virtual network having 600 ohms resistance image impedances has to be determined (and a network constructed if actual subjective determinations are to be made). Particularly difficult problems are encountered if the real connection contains no part in the section between JS and JR that has a 600 ohms image impedances (such as in a local call connection), but these can be overcome satisfactorily by calculation. Provided that a part is present having at least about 7 dB attenuation and 600 ohms image impedances, the problems can be overcome fairly easily.

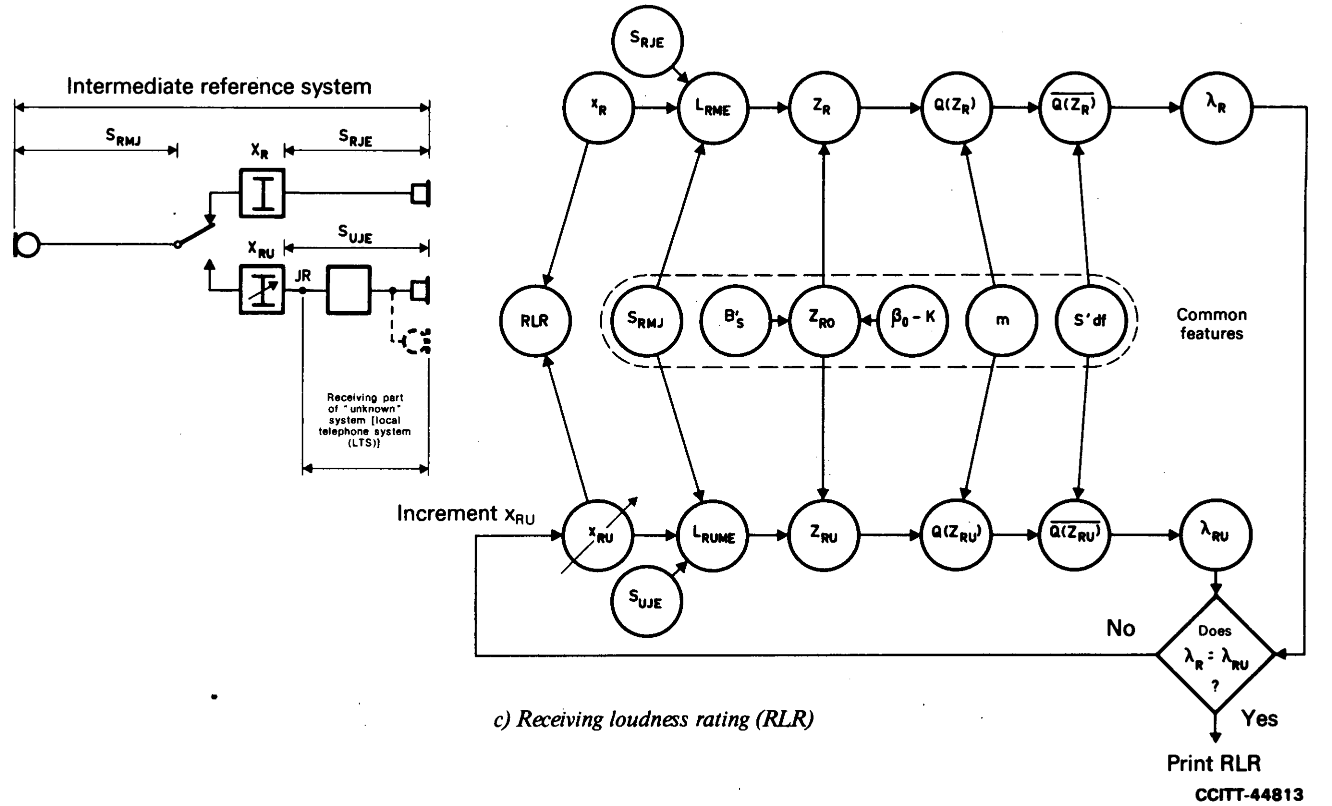
FIGURE 5/P.79

Flow diagrams illustrating determination of loudness ratings



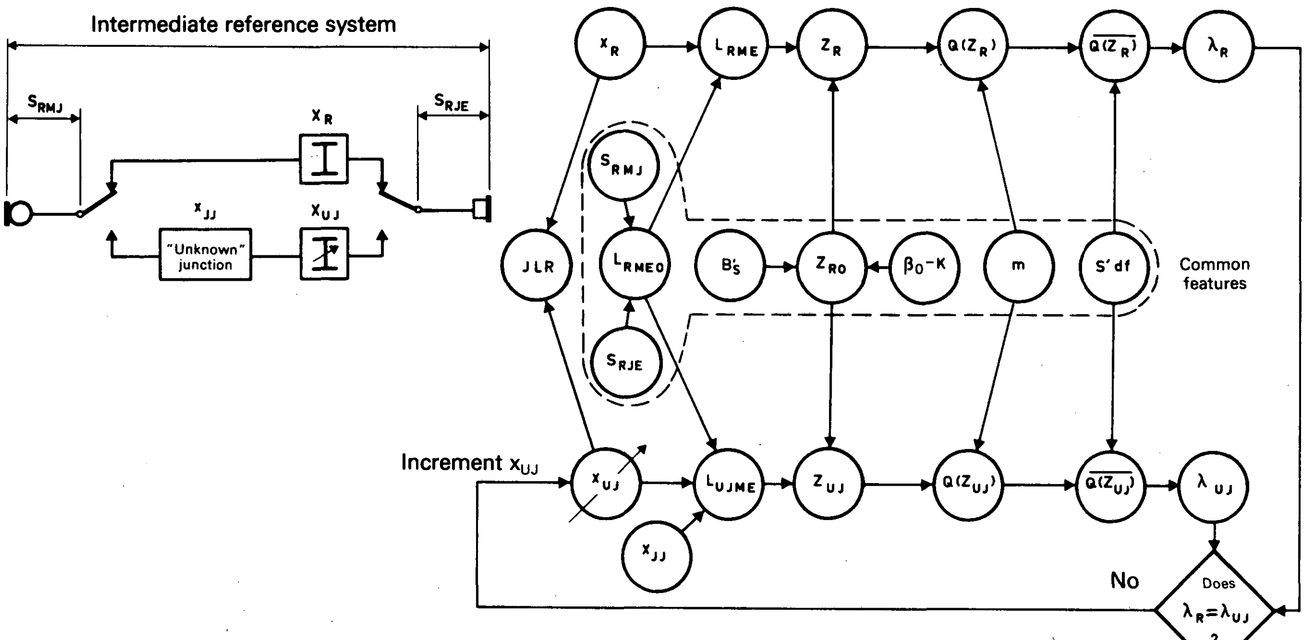
b) Sending loudness rating (SLR)

FIGURE 5/P.79 *(continued)*



c) Receiving loudness rating (RLR)

FIGURE 5/P.79 *(continued)*



d) Junction loudness rating (JLR)

CCITT-44822

FIGURE 5/P.79 *(end)*

Taking m as constant with the value 0.175, use can be made of the substitution:

$$W_i = -57.1 \log_{10} G\,\Delta f \tag{4-13}$$

Equation (4-11) can then be simplified in appearance to:

$$\overline{L_{UME}} = -57.1 \log_{10} \sum_i^N 10^{-(1/57.1)(L_{UME} + W_i)} \tag{4-14}$$

For the present purposes, the reference speech path will be taken as the "intermediate reference system" (IRS) defined in Recommendation P.48 and set with its attenuator at 0 dB; having fixed the reference speech path, L_{RME} becomes constant, i.e. independent of i. Therefore Equations (4-10) and (4-14) can be combined to form:

$$\text{loudness rating} = -57.1 \log_{10} \sum_i^N 10^{-(1/57.1)(L_{UME} - \overline{L_{RME}} + W_i)} \tag{4-15}$$

When rating commercial local telephone circuits, the values of L_{UME} can be obtained for any given "unknown" speech path combining appropriate sending and receiving sensitivites, S_{MJ} and S_{JE}, in appropriate combinations.

For determining an "overall loudness rating" (OLR),

$$L_{UME} = -(S_{UMJ} + S_{UJE}) \tag{4-16a}$$

For determining a sending loudness rating (SLR) of a local telephone circuit,

$$L_{URME} = -(S_{UMJ} + S_{RJE}) \tag{4-16b}$$

For determining a receiving loudness rating (RLR) of a local telephone circuit,

$$L_{RUME} = -(S_{RMJ} + S_{UJE}) \tag{4-16c}$$

and for determining a "junction" loudness rating (JLR)

$$\left.\begin{aligned} L_{UJME} &= -(S_{RMJ} + S_{RJE}) + x_{JJ} \\ &\text{and} \\ L_{RMEO} &= -(S_{RMJ} + S_{RJE}) \end{aligned}\right. \tag{4-16d}$$

Substituting these in Equation (4-15):

$$\mathrm{OLR} = -57.1 \log_{10} \sum_{i}^{N} 10^{(1/57.1)(S_{UMJ} + S_{UJE} + \overline{L_{RME}} - W_i)} \tag{4-17a}$$

$$\mathrm{SLR} = -57.1 \log_{10} \sum_{i}^{N} 10^{(1/57.1)(S_{UMJ} + S_{RJE} + \overline{L_{RME}} - W_i)} \tag{4-17b}$$

$$\mathrm{RLR} = -57.1 \log_{10} \sum_{i}^{N} 10^{(1/57.1)(S_{UJE} + S_{RMJ} + \overline{L_{RME}} - W_i)} \tag{4-17c}$$

$$\mathrm{JLR} = -57.1 \log_{10} \sum_{i}^{N} 10^{(1/57.1)(-x_{JJ} - L_{RMEO} + \overline{L_{RME}} - W_i)} \tag{4-18}$$

The terms $\overline{L_{RME}}$ and W_i are common to each of the Equations (4-17) and so further computational simplification is possible by making the following substitutions:

$$W_O = W_i - \overline{L_{RME}} \tag{4-18a}$$

$$W_S = W_i - S_{RJE} - \overline{L_{RME}} \tag{4-18b}$$

$$W_R = W_i - S_{RMJ} - \overline{L_{RME}} \tag{4-18c}$$

$$W_J = W_i + L_{RMEO} - \overline{L_{RME}} \tag{4-18d}$$

When the substitutions are made, the equations become:

$$\mathrm{OLR} = -57.1 \log_{10} \sum_{i}^{N} 10^{(1/57.1)(S_{UMJ} + S_{UJE} - W_O)} \tag{4-19a}$$

$$\mathrm{SLR} = -57.1 \log_{10} \sum_{i}^{N} 10^{(1/57.1)(S_{UMJ} - W_S)} \tag{4-19b}$$

$$\mathrm{RLR} = -57.1 \log_{10} \sum_{i}^{N} 10^{(1/57.1)(S_{UJE} - W_R)} \tag{4-19c}$$

$$\mathrm{JLR} = -57.1 \log_{10} \sum_{i}^{N} 10^{(1/57.1)(-x_{JJ} - W_J)} \tag{4-19d}$$

Table 2/P.79 shows the values for these "weighting" factors which have been derived from the information in Table 1/P.79 with $m = 0.175$.

TABLE 2/P.79

Weighting factors for calculating loudness ratings

Band No.	Mid-frequency (Hz)	Send W_s	Receive W_R	Junction W_J	Overall W_o
1	100	154.5	152.8	200.3	107.0
2	125	115.4	116.2	151.5	80.1
3	160	89.0	91.3	114.6	65.7
4	200	77.2	85.3	96.4	66.1
5	250	62.9	75.0	77.2	60.7
6	315	62.3	79.3	73.1	68.5
7	400	45.0	64.0	53.4	55.6
8	500	53.4	73.8	60.3	66.9
9	630	48.8	69.4	54.9	63.3
10	800	47.9	68.3	52.8	63.4
11	1000	50.4	69.0	54.1	65.3
12	1250	59.4	75.4	61.7	73.1
13	1600	57.0	70.7	57.6	70.1
14	2000	72.5	81.7	72.2	82.0
15	2500	72.9	76.8	71.1	78.6
16	3150	89.5	93.6	87.7	95.4
17	4000	117.3	114.1	154.5	76.9
18	5000	157.3	144.6	209.5	92.4
19	6300	172.2	165.8	245.8	92.2
20	8000	181.7	166.7	271.7	76.7

5.2 *Loudness rating calculations over a reduced bandwidth*

In practical cases the complete information for all 20 bands may not be available, or, for some extreme bands, may not be reliable. In such cases it will be desirable to restrict the frequency range over which calculations of loudness are made.

This may be done quite simply by using only those bands for which reliable figures exist and making an allowance equal to the loudness rating of the overall IRS connection calculated over the same reduced bandwidth. This allowance may conveniently be incorporated into the calculations by reducing the W weights uniformly by an appropriate figure, or by simply reducing (subtracting from) the resulting loudness rating by the allowance.

Table 3/P.79 gives some examples of the allowance to be applied for various reduced bandwidths.

Other allowances may be calculated by determining the IRS overall loudness rating for the required bandwidth.

TABLE 3/P.79

Allowance to be subtracted from *W* weights for reduced bandwidths

Bands	Allowance
3 – 18 (inclusive)	0.1 dB
3 – 17 (inclusive)	0.1 dB
4 – 18 (inclusive)	0.3 dB
4 – 17 (inclusive)	0.3 dB
6 – 16 (inclusive)	2.1 dB

Note – For loudness rating measuring instruments designed in accordance with the present Recommendation, the only bandwidth options recommended are:

i) 100-8000 Hz, bands 1-20 (inclusive)
ii) 200-4000 Hz, bands 4-17 (inclusive)

5.3 *Sending loudness rating (SLR) and receiving loudness rating (RLR) values to be used in the series G Recommendations*

Commercial measuring instruments complying with the present Recommendation use a band of 200 to 4000 Hz or even 100 to 8000 Hz. This is much wider than the band for which CCITT Recommendations specify an assured transmission (namely 300 to 3400 Hz). (See, for instance, Recommendations G.132 and G.151).

Thus, in a national system which may be included in an international connection, the loudness of the analogue telephone set should be considered as the inferior to the values measured herein.

It should also be noted that the loudness rating measurements of Recommendations P.64 to P.79 are to be made with a terminating impedance of 600 ohms. This is most often not the impedance appearing in the 2-wire part of the network. For various reasons, many Administrations now specify a complex nominal impedance. Thus, there will be a mismatch effect.

For SLR and RLR, an investigation has been made for a range of typical analogue telephone set sensitivity and impedance characteristics as well as nominal impedances. The result is that, with sufficient practical accuracy, 1 dB should be added to the measured values of SLR and RLR of *analogue* telephone sets in the LR planning of networks which can be included in an international connection. Thus, with the designation of SLR_w and RLR_w for the measure values:

$$SLR = SLR_w + 1$$

$$RLR = RLR_w + 1$$

The same correction, it should be noted, also applies when an unloaded subscriber cable is included in the mesurements of this Recommendation.

(When this correction is applied for planning, the effect of an unloaded subscriber's line on the LR is equal to its insertion loss at about 1 kHz. See also Annex A to Recommendation G.111.)

For digital sets, however, the correction is *not* needed because the codec and filters in the set limit the band to a certain extent.

As a rule it can be understood, from the context, when SLR and RLR values refer to planning or to measured (analogue) set values. However, when confusion might arise, it should be clearly stated whether the values refer to planning or measurements.

6 Sensitivity and transmission loss data required

The sending sensitivity of the local telephone system, S_{MJ}, should be determined in principle using real mouths and real speech but it is usually sufficient to make these measurements using an artificial mouth and suitable test signal. See Recommendation P.64 for particulars.

The receiving sensitivity of the local telephone system, S_{JE}, should be determined in principle using real ears. The determination of the sensitivity denoted by S_{Je}, using an artificial ear, is explained in Recommendation P.64 but this quantity differs from the quantity required here by the artificial/real ear correction L_E, that is:

$$S_{JE} = S_{Je} - L_E$$

The value of L_E usually depends upon the frequency and upon the manner in which the earphone is held to the ear.

Table 4/P.79 shows values obtained for one type of telephone held fairly closely to the ear. Use of these values for calculation has given reasonably good agreement with receiving and junction loudness ratings determined by subjective measurements in the CCITT Laboratory. Such calculations have used these values of L_E for both the IRS and the "unknown".

The values of S_{RJE} used to determine the values of W_s in Table 2/P.79 include a correction for L_E corresponding to the values of Table 4/P.79. The values of S_{UJE} used in the calculation defined by Equations (4-19a) and (4-19c) should also include a correction for L_E, using either the values of Table 4/P.79 or other values which might be considered more appropriate for the conditions of use.

Note that the values of L_E used for the IRS have some effect on the calculated values of junction loudness rating. This matter is receiving further study under Questions 8/XII [3] and 12/XII [4].

TABLE 4/P.79

Values of L_E

Frequency (Hz)	L_E (dB)	Frequency (Hz)	L_E (dB)
100	20.0	1000	−2.3
125	16.5	1250	−1.2
160	12.5	1600	−0.1
200	8.4	2000	3.6
250	4.9	2500	7.4
315	1.0	3150	6.7
400	−0.7	4000	8.8
500	−2.2	5000	10.0
630	−2.6	6300	12.5
800	−3.2	8000	15.0

The transmission loss x_{JJ} is the insertion loss between 600-ohms terminations of the chain of transmission elements between JS and JR in Figure 1/P.79. Direct summation (with due respect to sign) of this quantity with S_{UMJ} and S_{UJE} will not, in general, give L_{UME} exactly because there are usually some impedance mismatches. Care must therefore be taken to determine L_{UME} correctly when calculating overall loudness ratings. The inaccuracy will be severe when the transmission loss x_{JJ} is small and when the image impedances of the elements between JS and JR depart considerably from 600 ohms. The correct values for L_{UME} can be obtained by direct measurement or by calculation taking all impedance mismatches properly into account.

7 Restrictions of use

The calculation procedure described here and the values given for the parameters are suitable for calculating sending, receiving and junction loudness ratings. They may also be used for calculating overall loudness ratings and loudness insertion loss provided the complete speech paths concerned are restricted to the telephone frequency band, i.e. nominally to the range 300-3400 Hz.

They are not suitable for making comparisons between speech paths having considerable differences in frequency band.

The values of the parameters have been chosen to give reasonably good agreement with subjective loudness rating determinations by the CCITT Laboratory using the method described in Recommendation P.78. The most important utilization of Recommendation P.79 is a universally accepted method for calculating the electro-acoustic performance of telephone sets. However, Recommendation P.79 represents only with limited accuracy the speech and hearing characteristics of "ordinary people". This fact should be borne in mind if a detailed circuit loudness analysis is attempted for a telephone system. For further information, see Supplement No. 19.

8 Calculation of sidetone masking rating (STMR)

8.1 *Calculation from first principles*

Recommendation P.76 describes the principles underlying the sidetone masking rating method in which the human sidetone signal L_{MEHS} is treated as a masking threshold against which the telephone sidetone path loss, L_{meST}, is rated. As previously reported the human sidetone path loss, L_{MEHS}, has been determined [5] and is shown graphically in Figure 4/P.76, and in tabular form below in Table 5/P.79. Two sets of values are given in Table 5/P.79 for use depending on whether the conditions of interest are for an earphone coupling that is sealed (column 9) or with a typical leak included (column 10).

The calculation method for STMR makes use of the same underlying principles as described for sending and receiving loudness ratings in §§ 3 and 4. The calculation procedure is summarized by the expression:

$$\text{STMR} = \frac{10}{m} \log_{10} \frac{\sum 10^{\frac{mZ_l + 10 \log_{10} S' \Delta f}{10}}}{\sum 10^{\frac{mZ + 10 \log_{10} S' \Delta f}{10}}} \tag{8-1}$$

where

$$Z = B'_S - L_{meST} - L_E - 10 \log_{10} \left(10^{\frac{\beta_0 - K}{10}} + 10^{\frac{B'_S - L_{MEHS}}{10}}\right) \tag{8-2}$$

and

$$Z_l = B'_S + S_{RmJ} + S_{RJe} - L_E - 10 \log_{10} \left(10^{\frac{\beta_0 - K}{10}} + 10^{\frac{B'_S - L_{MEHS}}{10}}\right) \tag{8-3}$$

where the quantities used are as defined in earlier sections but where, for m, an index:

$$m = 0.225$$

The summations are normally extended over the range 100 Hz to 8 kHz but may be restricted if L_{meST} cannot be satisfactorily determined over the full bandwidth.

Table 5/P.79 lists the values for each of the quantities at the ISO frequencies.

TABLE 5/P.79

Listing of quantities necessary for the calculation of STMR

Band No.	f	B'_S	$\beta_0 - K$	$10 \log_{10} S' \Delta f$	IRS S_{RmJ}	IRS S_{RJe}	L_E	L_{MEHS}	
	Hz	dB 1 pW/m²/Hz	dB 1 pW/m²/Hz	dB	dB 1 V/Pa	dB 1 Pa/V	dB	dB Sealed	dB Un-sealed
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	100	57.3	17.5	−19.7	−45.8	−27.5	20	−2.7	11.6
2	125	60.2	14.4	−18.8	−36.1	−18.8	16.5	−4	10.6
3	160	62.0	10	−17.8	−25.6	−10.8	12.5	−5.4	7.1
4	200	63.0	5	−17	−19.2	− 2.7	8.4	−2.7	7.6
5	250	63.0	2.5	−16	−14.3	2.7	4.9	−2.8	7.4
6	315	62.4	− 0.4	−15.1	−10.8	7.2	1.0	−2.6	6.1
7	400	61.1	− 3	−14.4	− 8.4	9.9	−0.7	−0.7	3.5
8	500	59.3	− 5	−13.6	− 6.9	11.3	−2.2	5	5.7
9	630	57.0	− 6.3	−13.3	− 6.1	11.9	−2.6	13.2	8.9
10	800	54.4	− 8	−12.8	− 4.9	12.3	−3.2	19.9	16.2
11	1000	51.5	− 9	−12.4	− 3.7	12.6	−2.3	26.1	23.8
12	1250	48.4	− 8.5	−12.2	− 2.3	12.5	−1.2	23.7	23.7
13	1600	45.4	− 8	−11.9	− 0.6	13	−0.1	22	22
14	2000	42.3	− 9	−11.9	0.3	13.1	3.6	21.1	21.1
15	2500	39.5	−11.5	−12	1.8	13.1	7.4	22.1	22.1
16	3150	36.8	−13.8	−12.1	1.8	12.6	6.7	23.3	23.3
17	4000	34.6	−13	−12.4	−37.2	−31.6	8.8	24.2	24.2
18	5000	32.8	−12.5	−12.5	−52.2	−54.9	10.0	(26)	(26)
19	6300	31.5	−11.1	−13	−73.6	−67.5	12.5	(28)	(28)
20	8000	30.9	− 9	−14	−90	−90	15.0	(30)	(30)

8.2 *Calculation of STMR using W weights*

In § 4 above, the fundamental principles underlying the loudness rating procedure for sending, receiving, overall and junction loudness ratings were further developed, and a simplified equation derived which makes use of the W weights listed in Table 2/P.79 together with simplified equations (4-19a) to (4-19d). The equations (8-1), (8-2) and (8-3), applying to the STMR calculation, may also be reduced to a simplified equation that makes use of a set of W weights and a value of m unique to STMR, thus:

$$\text{STMR} = -\frac{10}{m} \log_{10} \sum_{1}^{N} 10^{(m/10)(-L_{meST} - L_E - W_M)} \qquad (8\text{-}4)$$

or, if sidetone sensitivities have been measured:

$$\text{STMR} = -\frac{10}{m} \log_{10} \sum_{M=1}^{N} 10^{(m/10)(S_{meST} - L_E - W_M)} \qquad (8\text{-}5)$$

where $m = 0.225$ and W_M take the values given in Table 6/P.79.

In deriving W weights for the unsealed condition (column 3, Table 6/P.79), values of L_E in accordance with column 8, Table 5/P.79 have been assumed for the reference path (IRS). When calculating STMR unsealed, appropriate values of L_E should be added to the L_{meST} values and inserted in the formula as indicated. In many cases the L_E values of column 8, Table 5/P.79 will be satisfactory.

For the sealed condition the weights of column 2, Table 6/P.79 should be used and the L_E values associated with L_{meST}, set to zero.

TABLE 6/P.79

Weighting factors for calculating STMR

Band No.	W_{MS} sealed	W_{ML} unsealed
(1)	(2)	(3)
1	110.4	94.0
2	107.7	91.0
3	104.6	90.1
4	98.4	86.0
5	94.0	81.8
6	89.8	79.1
7	84.8	78.5
8	75.5	72.8
9	66.0	68.3
10	57.1	58.7
11	49.1	49.4
12	50.6	48.6
13	51.0	48.9
14	51.9	49.8
15	51.3	49.3
16	50.6	48.5
17	51.0	49.0
18	49.7	47.7
19	50.0	48.0
20	52.8	50.7

8.3 *Comments on sealed versus unsealed conditions for the calculation of STMR*

In deriving values of L_{MEHS} for the sealed ear, very stringent measures were taken to eliminate leaks between the earcap of the test receiver and the subjects' ears. For L_{MEHS} unsealed a particular value of L_E was acoustically inserted at the receiver. The difference between the L_{MEHS} sealed and L_{MEHS} with leak can be seen by comparing columns 9 and 10 of Table 5/P.79. Over the most important parts of the frequency range this difference approximates to the value of L_E used at the receiver. In practice, rating differences (sealed-unsealed) are generally less than 1 dB.

This suggests that in practice any leak present will affect L_{MEHS} and L_{MEST} approximately equally, at least over a practical range of acoustic leaks. This in turn suggests that the L_{MEHS} will always have approximately the same masking effect with respect to L_{MEST} irrespective of any leak present and that for purposes of rating sidetone loudness STMR is expected to give better correlation with subjective effects if calculated for sealed ear conditions.

Use of the sealed condition is preferred, but Administrations may continue to use STMR unsealed for experimental purposes or where accumulation of data makes it sensible to do so, e.g. for certain existing specifications. If this is the case it must be clearly stated in the related documentation.

8.4 *Calculation of LSTR using W weights*

Listener sidetone rating is calculated using the same algorithm as STMR (Equation (8-5)) but the sidetone sensitivity used is that derived using a room noise source (see Recommendation P.64, § 9). Thus:

$$\mathrm{LSTR} = -\frac{10}{m}\log_{10}\sum_{M=1}^{N} 10^{(m/10)(S_{RNST} - L_E - W_M)} \tag{8-6}$$

where $m = 0.225$ and W_m take the values given in Table 6/P.79.

LSTR may also be calculated by using a value of S_{RNST} that has been determined by correcting S_{meST} by Δ_{Sm} (see Recommendation P.10, Recommendation P.65 § 9 and the Handbook on Telephonometry, § 3.3.17c), thus:

$$S_{RNST} \cong S_{meST} + \Delta_{Sm}.$$

If this method is chosen, the sidetone sensitivity S_{meST} should also have been determined using a wideband noise source.

Annex A to Recommendation G.111 describes a method applicable to transmission planning in which LSTR is determined by a STMR corrected by a weighted value of Δ_{SM}.

Information on other aspects of sidetone will be found in [6] and in the Annex to Question 9/XII [7], in Recommendations G.121 and P.11, and also in Supplement No. 11 at the end of this Volume.

References

[1] CCITT – Question 19/XII, Contribution COM XII-No. 1, Study Period 1985-1988, Geneva, 1985.

[2] CCITT – Question 7/XII, Contribution COM XII-No. 1, Study Period 1985-1988, Geneva, 1985.

[3] CCITT – Question 8/XII, Contribution COM XII-No. 1, Study Period 1985-1988, Geneva, 1985.

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

[5] CCITT – Contribution COM XII-No. 228/AP VII-No.115, Study Period 1977-1980, Geneva, 1980.

[6] CCITT – Question 9/XII, Contribution COM XII-234, Study Period 1981-1984, Geneva, 1984.

[7] CCITT – Question 9/XII, Contribution COM XII-1, Study period 1989-1992, Geneva 1988.

SECTION 7

SUBJECTIVE OPINION TESTS

Recommendation P.80 [1)]

METHODS FOR SUBJECTIVE DETERMINATION OF TRANSMISSION QUALITY

1 Introduction

This Recommendation contains advice to Administrations on conducting subjective tests in their own laboratories. The tests carried out in the CCITT Laboratory by using reference systems are described in Section 3 of this Volume.

In the course of developing items of telephone equipment, it is necessary to conduct various kinds of specialized tests to diagnose faults and shortcomings; such tests dedicated to the study of specific aspect of transmission quality are not discussed here. The present purpose is to indicate methods that have been found suitable for determining how satisfactory given telephone connections may be expected to be if offered as such for use by the public.

The methods indicated here are intended to be generally applicable whatever the form of any degrading factors present. Examples of degrading factors include transmission loss (often frequency dependent), circuit and room noise, sidetone, talker echo, nonlinear distortion of various kinds, propagation time, deleterious affects of voice-operated devices and changes in characteristics of telephone sets, including loudspeaking sets. Combinations of two or more of such factors have to be catered for.

2 Recommended methods

To be applicable for such a wide range of types of degrading factor given in § 1, the assessment method must reproduce as far as possible all the relevant features present when customers converse over telephone connections. Suitable methods are referred to as "Conversation Tests" and detailed prescriptions on the conduct of such tests as carried out by British Telecom are given in Supplement No. 2 at the end of this volume.

If the rather large amount of effort needed is available and the importance of the study warrants, transmission quality can be determined by service observations and recommended ways of performing these, including the questions to be asked when interviewing customers, are given in Recommendation P.82.

A disadvantage of the service observation method for many purposes is that little control is possible over the detailed characteristics of the telephone connections being tested. A method that largely overcomes this disadvantage but retains many of the advantages of service observations is that used by the AT&T Co. and termed SIBYL (refer to Supplement No. 5, Volume V, *Red Book*). According to this method, members of the staff of Bell Laboratories volunteer to allow a small proportion of their ordinary internal calls to be passed through special arrangements which modify the normal quality of transmission according to a test programme. If a particular call has been so treated the volunteer is asked to vote by dialling one of a set of digits to indicate his opinion. In this way all results are recorded by the controlling computer and complete privacy is retained.

---

1) This Recommendation was numbered P.74 in the *Red Book*.

3 Supplementary methods

Under certain conditions, it is permissible to dispense with the full conversation method and to use one-way listening-only tests. Suitable conditions apply for using a listening test when the degrading factor(s) under study affect the subjects only in their listening role. Attenuation/frequency distortion and nonlinear distortion caused by quantizing have been studied successfully by listening tests but it would be unwise to study the effects of sidetone, for example, by this method. Listening-only tests may also be misleading when assessing the effects of a factor, like circuit noise, when the magnitude of the degradation caused is substantial. In any case, sufficient comparison with the results from full conversation tests should be made before the results from listening-only tests are accepted as reliable.

Recommendation P.81 [1)]

MODULATED NOISE REFERENCE UNIT (MNRU)

(Malaga-Torremolinos, 1984; amended Melbourne, 1988)

The CCITT,

considering

(a) that the use of digital processes (64 kbit/s PCM A-law or μ-law, A/D/A encoder pairs, A/μ-law or μ/A-law converters, digital pads based on 8-bit PCM words, 32 kbit/s ADPCM, etc.) in the international telephone network has grown rapidly over the past several years, and this growth is expected to continue;

(b) that new digital processes are being standardized, e.g. 64 kbit/s 7 kHz wideband ADPCM;

(c) that there is a need for standard tools to measure the quantization distortion performance of digital processes [for example, 32 kbit/s ADPCM (Recommendation G.721) and 64 kbit/s 7 kHz wideband codec (Recommendation G.722)], so that the tools can be used for estimating the subjective transmission performance of international connections containing digital processes;

(d) that an objective speech quality assessment method has not yet been established;

(e) that, at the present time, subjective tests incorporating reference system conditions represent the only suitable method for measuring the speech transmission performance of digital processes;

(f) that expressing results in terms of a common reference system may facilitate comparison of subjective test results obtained at different laboratories,

recommends

(1) the use of a narrow-band Modulated Noise Reference Unit (MNRU) as the reference system in terms of which subjective performance of telephone bandwidth digital processes should be expressed;

(2) the use of a wideband MNRU as the reference system in terms of which subjective performance of wideband digital processes should be expressed [2)].

Note 1 – The MNRU can be realized using laboratory equipment or by computer simulation. Further information on the MNRU is given in the references listed at the end of this Recommendation.

---

1) This Recommendation was numbered P.70 in the *Red Book*

2) This specification is subject to future enhancement and therefore should be regarded as provisional.

Note 2 – The listening-only method presently proposed when using the MNRU in subjective tests is described in Supplement No. 14 at the end of this volume. See Recommendation P.80, § 3, for precautions concerning the use of listening-only tests.

Note 3 – Objective measurement methods which suitably reflect subjective quantization distortion performance of various types of digital processes do not exist at present. (For example, the objective technique of Recommendation G.712, based on sine-wave and band-limited noise measurements, are designed for PCM and do not measure appropriately the distortion induced by other systems such as ADPCM.) The artificial voice described in Recommendation P.50 may be relevant. Even if an objective method is developed, subjective tests will be required to establish correlation of subjective results/objective results for particular digital process types.

Note 4 – The wideband MNRU without noise shaping as described in this Recommendation is recommended[3]. Administrations are asked to comment on the need for a filter in the noise path after the multiplier (see Supplement No. 15), to shape the correlated noise spectrum. Some Administrations suggest the use of such a filter while others do not.

1 Introduction

The MNRU was originally devised to produce distortion subjectively similar to that produced by logarithmically companded PCM systems [1]. This approach was based on the views:

1) that network planning would require extensive subjective tests to enable evaluation of PCM system performance over a range of compandor characteristics, at various signal levels and in combination with various other transmission impairments (e.g. loss, idle circuit noise, etc.) at various levels, and

2) that it would be as reliable and easier to define a reference distortion system, itself providing distortion perceptually similar to that of PCM systems, in terms of which the performance of PCM systems could be expressed. This requires extensive subjective evaluation of the reference system when inserted in one or more simulated telephone connections, but leads to the possibility of simplified subjective evaluation of new digital processing techniques.

Various organizations (Administrations, scientific/industrial organizations), as well as the CCITT itself, have made extensive use of the MNRU concept for evaluating the subjective performance of digital processes (in arriving at Recommendations G.721 and G.722, for example). A modified version for use in evaluating codecs of wider bandwidth (70-7000 Hz) is now common practice. However, the actual devices used, while based on common principles, may have differed in detail, and hence the subjective results obtained may also have differed. (Differences in subjective testing methodology are also relevant.) The purpose of this Recommendation is to define the narrow-band and wideband versions of the MNRU as completely and in as much detail as possible in order to minimize the effects of the device, and of its objective calibration procedures, on subjective-test results.

2 General description

Simplified arrangements of the MNRU are shown in Figure 1a/P.81 for the narrow-band version and Figure 1b/P.81 for the wideband version. Speech signals entering from the left are split between 2 paths, a signal path and a noise path. The signal path provides an undistorted (except for bandpass filtering) speech signal at the output. In the noise path, the speech signal instantaneously controls a multiplier with an applied gaussian noise "carrier" which has a uniform spectrum between 0 Hz and a frequency at least twice the cutoff frequency of the lowpass portion of the bandpass filter. The output of the multiplier consisting of the noise modulated by the speech signal, is then added to the speech signal to produce the distorted signal.

The attenuators and switches in the signal and noise paths allow independent adjustment of the speech and noise signal levels at the output. Typically, the system is so calibrated that the setting of the attenuator (in dB) in the noise path represents the ratio of instantaneous speech power to noise power, when both are measured at the output of the band-pass filter (terminal OT).

[3] This specification is subject to future enhancement and therefore should be regarded as provisional.

For this Recommendation, the decibel representation of the ratio is called Q_N for the narrow-band version and Q_W for the wideband version.

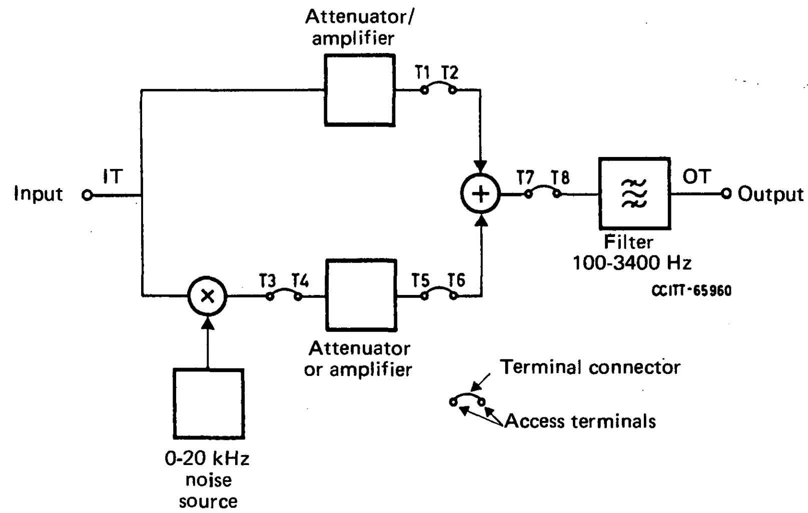


FIGURE 1a/P.81

Basic arrangement of narrow-band MNRU

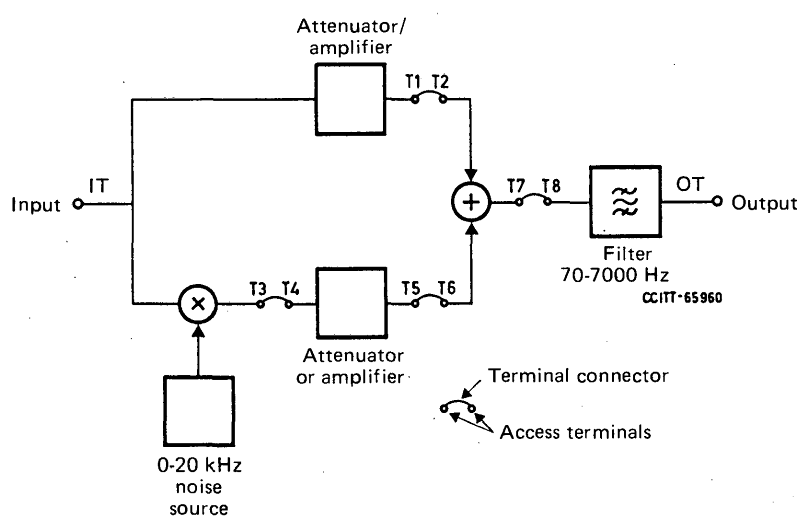


FIGURE 1b/P.81

Basic arrangement of wideband MNRU

3 Performance specifications

3.1 *General*

The specifications in this section apply both to hardware implementations and software simulations.

For practical implementations, the actual signal levels and noise levels may be increased or decreased to meet special needs. In such cases, the level requirements detailed below will have to be modified accordingly.

3.2 *Signal path*

The requirements under this heading refer to the MNRU with infinite attenuation in the noise path of Figures 1a/P.81 and 1b/P.81; separate resistive terminations at the terminals T5 and T6 (unlinked) will achieve this.

The frequency response of the signal path (i.e. between terminals IT and OT of Figures 1a/P.81 and 1b/P.81) should be within the limits of Figure 2a/P.81 for the circuit of Figure 1a/P.81 and within the limits of Figure 2b/P.81 for the circuit of Figure 1b/P.81.

The loss between terminals IT and OT for a 0 dBm, 1 kHz input sine wave should be 0 dB. Over the input level range +10 dBm to −50 dBm, the loss should be 0 dB ± 0.1 dB.

Any harmonic component should be at least 50 dB below the fundamental at the system output (terminal OT in Figures 1a/P.81 and 1b/P.81) for any fundamental frequency between 125 Hz and 3000 Hz in a narrow-band system and 100 Hz and 6000 Hz in a wideband system.

The idle noise generated in the signal path must be less than −60 dBm, measured at terminal OT, in order to conform with § 3.4.

It is recommended that the level of speech signals applied to the terminals IT should be less than −10 dBm (mean power while active, i.e. mean active level according to Recommendation P.56) in order to avoid amplifier peak-clippings of the signal, and be greater than −30 dBm to ensure sufficient speech signal-to-noise ratio.

3.3 *Noise path*

The requirements under this heading refer to the MNRU with infinite attenuation inserted into the signal path of Figures 1a/P.81 and 1b/P.81; separate resistive terminations at the terminals T1 and T2 (unlinked) will achieve this.

3.3.1 *Linearity as function of input level*

With a Q_N setting of 0 dB in the circuit of Figure 1a/P.81, or a Q_W setting of 0 dB in the circuit of Figure 1b/P.81, as the case may be, the noise level at the system output (terminal OT) should be numerically equal to the sine wave level at the input terminal (terminal IT). A correspondence within ± 0.5 dB should be obtained for input levels from +5 dBm to −45 dBm, and for input frequencies from 125 Hz to 3000 Hz in a narrow-band system and 100 Hz to 6000 Hz in a wideband system.

3.3.2 *Noise spectrum*

For a narrow-band system, when Q_N is set to 0 dB, input sine waves applied to terminal IT in Figure 1a/P.81 with levels from +5 to −45 dBm and frequencies from 125 Hz to 3000 Hz should result in a flat noise system spectrum density at the output of the multiplication device (terminal T3 of Figure 1a/P.81) within ± 1 dB over the frequency range 75 Hz to 5000 Hz. The spectrum density should be measured with a bandwidth resolution of maximum 50 Hz.

For a wideband system, when Q_W is set to 0 dB, input sine waves applied to terminal IT in Figure 1b/P.81 with levels from +5 to −45 dBm and frequencies from 100 Hz to 6000 Hz should result in a flat noise system spectrum density at the output of the multiplication device (terminal T3 of Figure 1b/P.81) within ± 1 dB over the frequency range 75 Hz to 10 000 Hz. The spectrum density should be measured with a bandwith resolution of maximum 50 Hz.

3.3.3 *Amplitude distribution*

The amplitude distribution of the noise at the system output should be approximately gaussian.

Note – A noise source consisting of a gaussian nose generator followed by a peak clipper with a flat spectrum from near zero to 20 kHz will produce a satisfactory output noise at terminal OT.

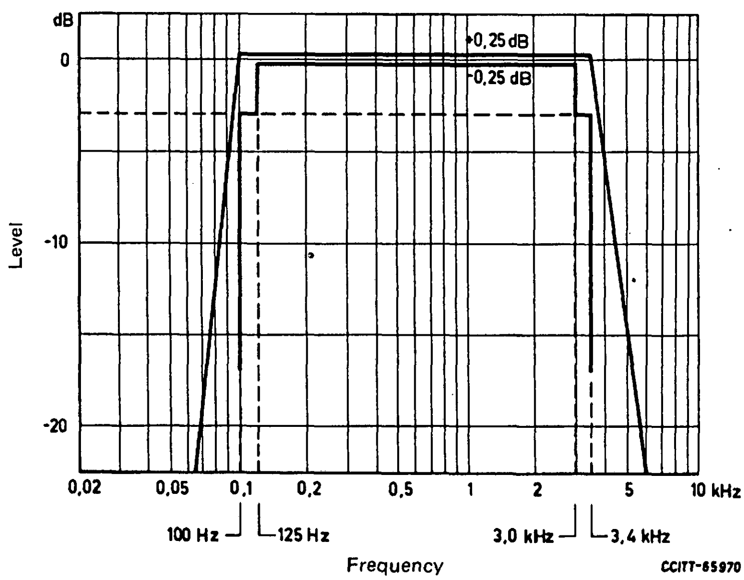


FIGURE 2a/P.81

Requirements for output filter of the narrow-band MNRU

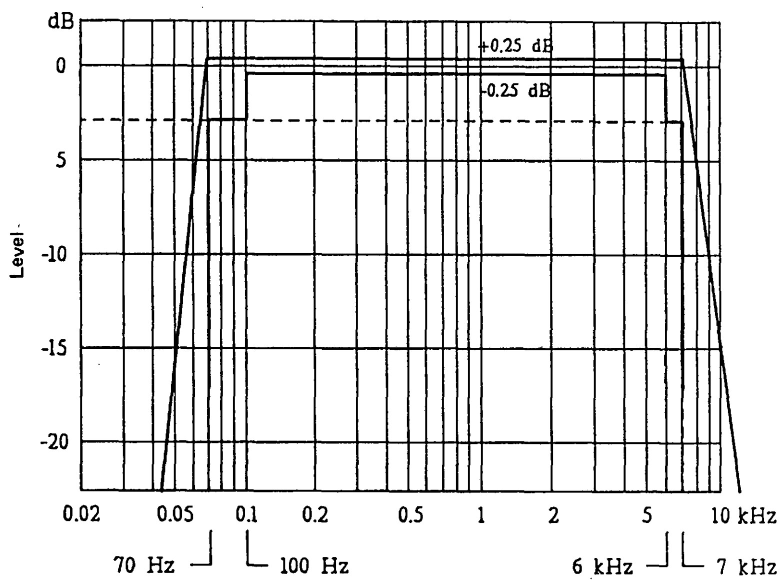


FIGURE 2b/P.81

Requirements for output filter of the wideband MNRU

3.3.4 *Noise attenuators*

The loss of the noise attenuator(s) i.e. between terminals T4 and T5 in Figures 1a/P.81 and 1b/P.81, should be within ± 0.1 dB of the nominal setting. The attenuator(s) should at least allow Q_N and Q_W settings in the range −5 dB to 45 dB, i.e. a 50 dB range.

3.4 *Combined path*

The requirements under this heading refer to the MNRU with both speech and noise paths simultaneously in operation.

With Q_N or Q_W (as the case may be) set to zero, and the input terminated by an equivalent resistance, the idle noise generated in the combined path should be less than −60 dBm when measured at the system output (terminal OT).

References

[1] LAW (H. B.), SEYMOUR (R. A.): A reference distortion system using modulated noise, *The Institute of Electrical Engineers*, pp. 484-485, November 1962.

Bibliography

CCITT – Contribution COM XII-No. 63, *Some considerations on specifications for modulated noise reference unit*, NTT, Japan, Study Period 1981-1984.

CCITT – Contribution COM XII-No. R4, pp. 71-79, Study Period 1981-1984.

CCITT – Contribution COM XII-No. 119, *Description and method of use of the modulated noise reference unit (MNRU/MALT)*, France, Study Period 1981-1984.

Recommendation P.82 [1)]

METHOD FOR EVALUATION OF SERVICE FROM THE STANDPOINT OF SPEECH TRANSMISSION QUALITY

(Geneva, 1976; amended at Malaga-Torremolinos, 1984)

1 General

The CCITT recommends that Administrations make use of telephone users' surveys in the manner of Recommendation E.125 [1] as a means of measuring speech transmission quality on international calls.

Such surveys being call-related (in this instance to the last international call made) can be conducted either by the full use of the Recommendation E.125 questionnaires (where other valuable information is obtained on users' difficulties, e.g. knowing how to make the call, difficulties in dialling or understanding tones, etc.) or by making use of those questions solely related to transmission quality which appear in Annex A.

Note – The evaluation of the transmission performance may be altered by difficulties in setting-up call. Hence the response to incomplete questionnaires should be considered with some reservation.

[1)] This Recommendation was numbered P.77 in the *Red Book*

2 Conduct of surveys

In order to make valid comparisons between data collected in different countries, Recommendation E.125 should be strictly adhered to. Specifically the preamble to the Recommendation, the notes of intended use of the questionnaires and the precise order and wording of the questions should be rigidly followed. In some cases, however, an exception will be made and Question 10.0 will be replaced by the wording indicated in Annex B (detailed information is given in [3]).

Note – This alternative version has the advantage of simplifying the classification of responses to open end probes by experts, as well as increasing the sensitivity to some types of impairments such as delay. These advantages should be weighed against the additional interview time which may be required.

3 Treatment of results

To provide quantitative information suitable for comparisons, the subjective assessments (e.g. those obtained from Question 9.0 of Annex A) of excellent, good, fair or poor (see Note) should be accorded scores of 4, 3, 2 and 1, respectively and a mean opinion score (MOS) calculated for all associated responses. Similarly for all those experiencing difficulty (under Question 10.0 of Annex A or, alternatively, Question 10.0 of Annex B) a percentage of the total responses should be calculated. These two criteria of MOS and percentage difficulty are now internationally recognized and have been measured under many different laboratory simulated connections and practical situations.

The results can be classified in a number of ways, e.g. in terms of the call-destination countries or by nature/composition of the connection i.e. cable/satellite circuits, presence or otherwise of echo suppressors etc. Typical methods of presentation of the results are shown in [2], in this case for several countries. It should be noted that in all presentations it is essential to show the number of responses.

Note – Among the reasons which lead to the limitation of users' opinions of transmission quality to four classes, i.e. excellent, good, fair and poor, is the following. The experience gained in human factor investigations has shown that when a question which requires a selection from several different classifications is posed in aural form, e.g. by face-to-face interview or by telephone as with Recommendation E.125, the respondent is frequently unable to carry a clear mental separation of more than four categories. As a consequence, he is unable to draw on his short-term memory and judgement ability in a sufficiently precise manner to avoid confusion and gives an unreliable response. This restriction does not apply to other situations where a written presentation of the choices is used, in which case frequently five or more classes may be appropriate and shown to yield reliable responses.

ANNEX A

(to Recommendation P.82)

Extract from the questionnaire annexed to Recommendation E.125

Reproduced below are the questions relating to transmission quality which appear in the questionnaire annexed to Recommendation E.125.

The CCITT recommends that this Annex should be used when customers' general impressions of transmission performance are required.

9.0 . . . .

Which of these four words comes closest to describing the quality of the connection during conversation?

9.1 – *excellent*

9.2 – *good*

9.3 – *fair*

9.4 – *poor*

1 ☐
2 ☐ } 48
3 ☐
4 ☐

10.0	*Did you or the person you were talking to have difficulty in talking or hearing over that connection?*	YES 1 ☐ NO 2 ☐	49

(If answer is "yes") probe for nature of difficulty, but without suggesting possible types of difficulty, and copy down answers verbatim: e.g. *"Could you describe the difficulty a little more?"*

. . . .

. . . .

At end of interview, categorize the answers in terms of the items below:

10.1	– low volume	1 ☐	50
10.2	– noise or hum	1 ☐	51
10.3	– distortion	1 ☐	52
10.4	– variations in level, cutting on and off	1 ☐	53
10.5	– crosstalk	1 ☐	54
10.6	– echo	1 ☐	55
10.7	– complete cut off	1 ☐	56
10.8	– other (specify) . . . . ←	1 ☐	57

Note – Responses to Questions 10.1 to 10.8 are only obtained from customers who have expressed difficulty in Question 10.0.

ANNEX B

(to Recommendation P.82)

Alternative version for Question 10.0 of questionnaire annexed to Recommendation E.125

Studies at AT&T have shown that the verbatim responses describing impairments (requested after Question 10.0 of Annex A) are often too imprecisely worded to permit accurate classification by interviewers who are not experienced in transmission studies. A typical solution to this problem has been to convene a panel of experts to classify the responses, a method which may become impractical as the size and number of user reaction tests increases. This annex presents an alternative approach developed in 1976 and used widely since then by AT&T to measure customer's perceptions of transmission quality on domestic and international telephone connections. The approach involves a more complicated technique of probing for impairments which simplifies the ultimate task of classifying the responses. The alternative of Question 10.0 is reproduced below.

The CCITT recommends that this annex should be used for diagnostic purposes only.

10.0	*Did you have any difficulty talking or hearing over that connection?*	YES 1 ☐ NO 2 ☐	49

Do not probe: If the person volunteers an explanation, write it down. . . . .

. . . .

On question 10.1-10.8, attempt to read entire text before respondent replies.

10.1 *Now I'd like to ask some specific questions about the connection.*

If the person has already described difficulty, add:

(In view of what you've already said, some of these may seem repetitious, but please bear with me). First, during your conversation on that call, did you hear your own voice echoing back, or did your own voice sound hollow to you?

10.1.1	– echo hollow (own voice)	1 ☐	50
10.1.2	– neither	2 ☐	
10.1.3	– don't remember/not sure	3 ☐	
10.1.4	– other (specify) . . . . ←	4 ☐	

10.2 *Did you hear another telephone conversation on the telephone network at the same time as your own?*

10.2.1 – other conversation 1 ☐

10.2.2 – no 2 ☐

10.2.3 – don't remember/not sure 3 ☐

10.2.4 – other (specify) . . . . ← 4 ☐

51

10.3 *Now I'd like you to think about the voice of the person you were talking to. Was the volume of the voice low as if the person were faint and far away; did the voice fade in and out; or was the voice interrupted or chopped up at times?*

10.3.1 – low volume 1 ☐

10.3.2 – fading 2 ☐

10.3.3 – chopping 3 ☐

10.3.4 – none 4 ☐

10.3.5 – don't remember/not sure 5 ☐

10.3.6 – other (specify) . . . . ← 6 ☐

52

10.4 *How did the voice of the person your were talking to sound to you: did it echo or sound hollow and tinny; or did it sound fuzzy or unnatural?*

10.4.1 – echo, hollow 1 ☐

10.4.2 – fuzzy, unnatural 2 ☐

10.4.3 – none 3 ☐

10.4.4 – don't remember/not sure 4 ☐

10.4.5 – other (specify) . . . . ← 5 ☐

53

10.5 *Now let me describe three kinds of noise. Tell me if you noticed any of these noises during your conversaiton: a rushing or hissing sound; a frying and/or sizzling, crackling sound; or a humming or buzzing sound?*

10.5.1 – rushing, hissing 1 ☐

10.5.2 – frying and/or sizzling, cackling 2 ☐

10.5.3 – humming, buzzing 3 ☐

10.5.4 – none 4 ☐

10.5.5 – don't remember/not sure 5 ☐

10.5.6 – other (specify) . . . . ← 6 ☐

54

10.6 *Now let me describe three more kind of noise. Tell me if you noticed any of these during your conversation: a clicking sound; a series of musical tones or beeps; or a continuous high-pitched tone?*

10.6.1 – clicking 1 ☐

10.6.2 – tones or beeps 2 ☐

10.6.3 – high-pitched tone 3 ☐

10.6.4 – none 4 ☐

10.6.5 – don't remember/not sure 5 ☐

10.6.6 – other (specify) . . . . ← 6 ☐

55

10.7 *Did the other person seem slow to respond, as if there were delay or time lag in the conversation?*

10.7.1	– yes	1 ☐	56
10.7.2	– no	2 ☐	
10.7.3	– don't know	3 ☐	
10.7.4	– other (specify) . . . . ←	4 ☐	

10.8 *Would you please try to remember the background noise in the area around your telephone (e.g. noise from air-conditioning plant unit, road traffic, office equipment or other people talking) when you made the call. Which of the following categories best describes it?*

10.8.1	– very noisy	1 ☐	57
10.8.2	– noisy	2 ☐	
10.8.3	– quiet	3 ☐	
10.8.4	– very quiet	4 ☐	
10.8.5	– other (specify) . . . . ←	5 ☐	

10.9 *Which of the categories listed below best describes the extent to which you heard your own voice through your telephone when you were talking?*

10.9.1	– could not hear it	1 ☐	58
10.9.2	– could hear it now that you have drawn my attention to it	2 ☐	
10.9.3	– did notice it – not loud	3 ☐	
10.9.4	– did notice it – loud	4 ☐	
10.9.5	– other (specify) . . . . ←	5 ☐	

10.10 *Was there anything else about the connection you'd like to mention?*

Yes – What? (Write in)

. . . .

. . . .

. . . .

Coding instructions:	YES	NO	
– is there a written comment?	1 ☐	2 ☐	59
– does the comment apply to this call?	1 ☐	2 ☐	60
– does it mention an impairment?	1 ☐	2 ☐	61
– has it been mentioned already?	1 ☐	2 ☐	62
– other (specify)	1 ☐	2 ☐	63

Note – The responses to the specific questions are only obtained from customers who have expressed difficulty in Question 10.0. This may prevent the diagnosis of certain impairments (the bias produced is more serious than that mentionned at the end of Annex A).

References

[1] CCITT Recommendation *Inquiries among users of the international telephone service,* Red Book, Vol. II, Rec. E.125, ITU, Geneva, 1985.

[2] CCITT – Question 2/XII, Annex 2, Contribution COM XII-No. 1, Study Period 1977-1980, Geneva, 1977.

[3] CCITT – Question 2/XII, Annex, Contribution COM XII-No. 171, Study Period 1977-1980, Geneva, August 1979.

SUBJECTIVE LISTENING TEST METHOD FOR EVALUATING DIGITAL CIRCUIT MULTIPLICATION AND PACKETIZED VOICE SYSTEMS[1]

(Melbourne, 1988)

1 Introduction

1.1 *Purpose*

The purpose of this Recommendation is to describe a subjective listening test method which can be used to compare the performance of Digital Circuit Multiplication Equipment (DCME) and packetized voice systems.

Many of the degradations found in DCME or packetized voice systems have not been tested before and their effects on other systems in the network are unknown. Therefore the only definitive method is the conversation test where the effects of non-linearity, delay, echo, etc. and their interactions can be verified.

For DCME systems, degradations can include not only the effects of variable bit-rate coding, DSI gain (channel allocation), clipping, freezeout and noise contrast, but also those due to non-linearities in the speech detection system, such that the system may function differently for different speech input levels or activity factors. For packetized voice systems the subjective effect, for example, of "lost packets" is unknown.

Listening tests play an important preliminary role in the assessment, and can supply useful information serving to narrow the range of conditions needing a complete conversation test. Moreover, listening tests of the effects of the impairments produced by DCME, in association with an evaluation of the effects of delay added by the DCME, using the echo tolerance method described in Recommendation G.131, can give a good indication of the overall performance of such systems and allow reasonable comparisons to be made. In addition, the delay evaluation should determine whether or not the use of DCME in a network setting will require additional echo control. This listening test method will not provide results useful for generating network application rules based on factors such as the quantizing distortion unit (qdu). Future improvements of the test will allow such results to be obtained.

Evaluation of DCME in tandem with other DCME has not been considered at this stage nor have the effects of systems using encoding at different rates. This Recommendation will subsequently be updated when information on these specific points becomes available.

This Recommendation confines itself solely to listening tests; a separate Recommendation P.85, on conversation tests, will be formulated when sufficient information on evaluation techniques is available. Alternatively, this Recommendation may be revised to include conversation test methods.

1.2 *Definitions*

1.2.1 **digital circuit multiplication equipment (DCME)**

A general class of equipment which permits concentration of a number of 64 kbit/s PCM encoded input speech circuits onto a reduced number of transmission channels.

This equipment allows an increase in the circuit capacity of the system. The capacity of speech and voiceband data can both be increased by the use of DCME.

[1] The specifications in this Recommendation are subject to future enhancement and therefore should be regarded as provisional.

1.2.2 **digital circuit multiplication system** (DCMS)

A telecommunication system comprised of two or more DCME terminals connected by a digital transmission system providing a pool of bearer channels. The DCMS supports:

i) 64 kbit/s clear channels for ISDN services (can be used in the bearer pool),

ii) voiceband data (dial-up) up to and including 9600 bit/s V.29. Group III facsimile is also included under this heading,

iii) voice services in the frequency range 300-3400 Hz, carried at 56 or 64 kbit/s,

iv) 64 kbit/s clear (not ISDN dial-up),

v) sub-64 kbit/s digital data.

1.2.3 *Circuit versus packet mode*

Internally the DCME may employ a circuit or a packet mode for the transmission of speech or data. In the circuit mode, bearer channels are derived by providing suitable time slots on the transmission facility interconnecting the DCME terminal equipment. In the packet mode virtual bearer channels are created and the speech or data samples are put into one or more packets of fixed or variable length. The packets are addressed to the destination circuit and transmitted in a virtual channel on the transmission facility one at a time. Thus, in the circuit mode the transmission facility can be thought of as carrying a number of bearer channels multiplexed together, while in the packet mode the facility is thought of as a single high speed channel logically divided into virtual channels which transmits packets one at a time.

1.2.4 **single clique working (point-to-point operation)**

The system of two DCMEs interconnected by one set of bearer channels. This working of a DCME is the most efficient mode of operation for a DCMS. It utilizes the maximum bearer pool capacity and the minimum inter-DCME control information. It is an exclusive mode of operation. Another term for point-to-point is circuit-based DCMS. Figure 1/P.84 shows an example of point-to-point or circuit-based DCMS.

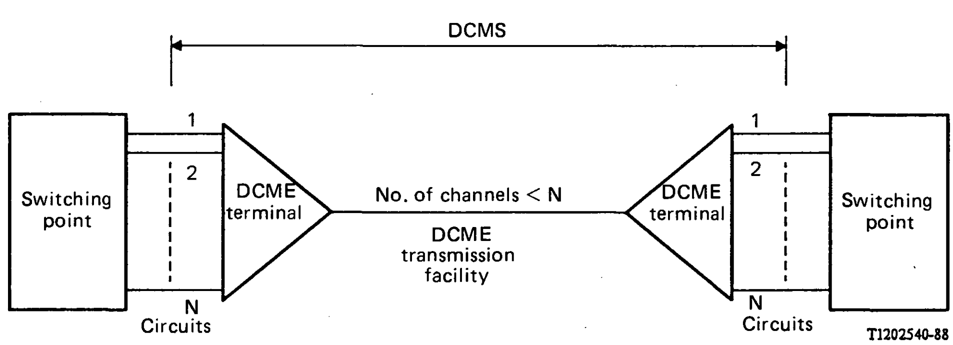


FIGURE 1/P.84

Point-to-point or circuit-based DCMS

1.2.5 **multi-clique working (point-to-multipoint operation)**

A single DCME working to more than one DCME each on a point-to-point destination basis; designations are split and are therefore not interactive. Multi-clique working reduces the traffic handling capacity compared with point-to-point operation, due to a reduction in bearer capacity. Single clique working is the equivalent of point-to-point operation.

1.2.6 **multi-destination operation**

Many DCMEs working over a common bearer capacity pool, enabling interactive working. This is the equivalent of a TDMA satellite system. Traffic handling capacity is drastically reduced since the bearer becomes very small, due to inter-DCME control messages and inter-terminal operation reducing the bearer capacity. Another term for multi-destination DCMS is network-based DCMS. Figure 2/P.84 shows an example of this.

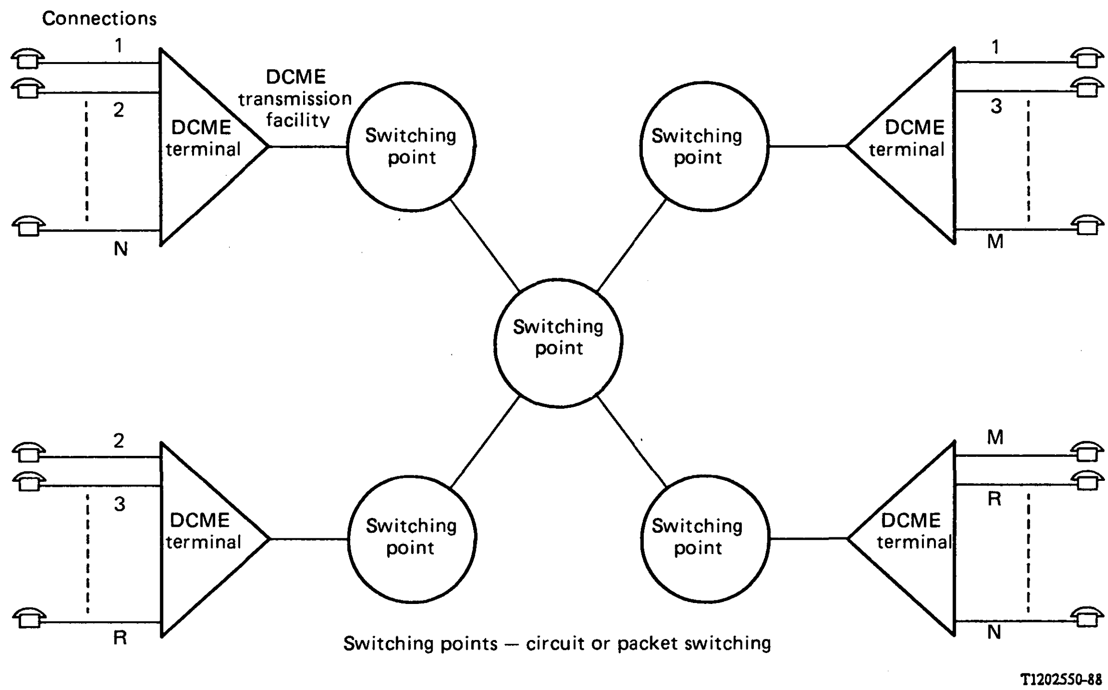


FIGURE 2/P.84

Multi-destination or connection-based DCMS

1.2.7 **low rate encoding (LRE)**

Speech coding methods with bit rates less than 64 kbit/s, e.g. the 32 kbit/s ADPCM transcoder, (Recommendation G.721). This is one technique commonly used in DCME to increase the circuit capacity.

1.2.8 **digital speech interpolation (DSI)**

This is a technique whereby advantage can be taken of the inactive periods during a conversation, creating extra channel capacity. Speech activity is typically 30-40%, on average, which can produce a DSI gain of up to 3 : 1, but generally in the range of 2 : 1 to 2,5 : 1.

1.2.9 **LRE gain, DSI gain, DCME gain**

LRE gain is the factor by which the 64 kbit/s rate of the incoming circuits is reduced when LRE is used for coding within the DCME. For example, when a transcoder conforming to Recommendation G.721 is used, the LRE gain will equal 2. The LRE gain is 1 when no transcoding is used.

DSI gain is the ratio of the number of active speech input circuits to the number of bearer channels used to transport this speech, where the same encoding rate is used for circuits and bearer channels. The DSI gain is constrained by the number of input circuits and the speech activity factor and other input speech characteristics. The DSI gain is 1 when DSI is not used.

The DCME gain is the product of the LRE and DSI gain factors.

1.2.10 **DCME overload**

The instant when the number of instantaneously active input circuits exceeds the maximum number of "normal" bearer channels available for DSI.

1.2.11 **freezeout**

The condition when an input circuit becomes active with speech and cannot be immediately assigned to a bearer channel, due to lack of availability of such channels.

1.2.12 **freezeout fraction**

The percentage of speech lost, obtained by averaging over all input circuits for a given time interval, e.g. one minute.

1.2.13 **transmission overload**

The condition when the freezeout fraction goes beyond the value set in accordance with the speech quality requirements.

1.2.14 **clipping**

An impairment occurring in DSI systems employing speech detectors whereby the detector, due to the time it takes to recognize that speech is present, can cut off ("clip") the start of the speech utterance. Competitive clipping is the impairment caused by the overload control strategy which allows freezeout to occur when bearer channels are temporarily unavailable. Another name for the competitive clipping overload control strategy is sample dropping.

1.2.15 **variable bit rate (VBR)**

An overload control strategy often used to cope with traffic peaks and hence freezeout problems. Temporary, additional bearer channels (overload channels) are created. Several VBR techniques are available:

i) Graceful overload is one technique to reduce the bit rate. For example, a 4-bit sample 32 kbit/s ADPCM channel can be reduced on demand to a minimum of a 3-bit sample 24 kbit/s, and the VBR will average across the DCMS somewhere between 3 and 4 bits. The dynamic load control (DLC) will operate when the predicted traffic loading rises above a preset VBR.

ii) Permanent 3-bit allocation set on block of channels. These channels operate solely in a 3-bit mode.

The different reduction techniques available have different subjective performances.

1.2.16 **queuing**

An overload control strategy employing buffer memory in the DCME transmitter to store speech samples while waiting for a bearer channel to become available.

1.2.17 **dynamic load control (DLC)**

An overload control strategy in which the DCMS signals to the associated switch that the traffic load the switch is generating, or is predicted to generate, cannot be transmitted satisfactorily by the DCMS and that the switch should reduce its demand on the DCMS by a holding signal sent to the circuits when they become idle.

1.2.18 **load carrying capacity**

The load carrying capacity is defined as the maximum offered speech load plus "overhead" load (see § 1.2.19) that the DCME channels can carry without forced loss of any speech samples. DCME overload is defined to occur when the instantaneously offered load exceeds the carrying capacity of the DCME bearer channels.

1.2.19 **applied and offerd load**

The applied load consists of the speech bursts entering the DCME on the active circuits. Thus, applied load is a function of the number of active circuits and the speech activity on the circuits.

The offered load consists of the applied load plus any additional load (overhead) generated by the DCME messages and control information. The offered load is the load presented to the DCME bearer channels. If the offered load is less than the load-carrying capacity of the channels, then all the offered load is carried by the DCME. However, if the offered load exceeds the capacity of the bearer channels, then, depending upon the overload strategy of the DCME, some of the offered load will be lost through competitive clipping (sample dropping). The DCME may employ variable bit rate coding so that, should the freezeout fraction exceed some preset limit, the DCME can momentarily increase the load-carrying capacity of the bearer channels (creation of overload channels) in order to accommodate the extra load. Dynamic load control may also be used to limit the applied load.

The instantaneous load is a function of the statistics of the input speech and the DCME overhead traffic, and is difficult to characterize mathematically. However, the long-term time average applied load can be calculated as follows:

$$L_a = N \frac{\alpha}{\alpha + \beta},$$

where L_a is the average applied load, α is the average speech burst length, β is the average silence length, and N is the number of circuits in use. The term $\alpha/(\alpha + \beta)$ is equal to the average speech activity. The applied load is measured at the input to the DCME on the circuits. Thus, the average load on the DCME can be externally controlled by varying the number of circuits in use, N, or the speech activity factor, $\alpha/(\alpha + \beta)$.

Similarly, average offered load is a useful concept, and it can be calculated from this formula:

$$L_o = N \frac{\alpha(k + 1)}{\alpha + \beta} + G,$$

where L_o is the average load offered to the bearer channels, the term k is a constant which accounts for the "stretching" effect that the speech detector has on the activity factor, and the term G is a load factor that accounts for the system overhead traffic (e.g. control messages). Thus, the average offered load, L_o, will almost always be larger than the average applied load, L_o.

1.3 *Test philosophy*

In order for a test to satisfactorily evaluate DCME performance the test methodology should meet certain conditions. These are as follows:

i) the method should use principles, procedures, and instrumentation that are acceptable to CCITT;

ii) the method should be adaptable to different languages and should yield results that are comparable to previous test results;

iii) the method should permit DCME performance to be compared subjectively (or objectively) to reference conditions. Examples of suitable reference conditions are hypothetical reference connections (HRCs), white noise and speech correlated noise. The HRCs should model the facilities the DCME is designed to replace, when these facilities are known. The results of the comparisons should permit making "equivalence statements" about the DCME, e.g. a DCME system is subjectively equivalent to x asynchronously tandemed 64 kbit/s PCM systems. Ideally, the method should yield results from which a network application rule can be derived;

iv) the DCME should be tested with a realistic traffic load simulator and circuit-under-test signal conditions applied. Most of the transitory impairments arise when the DCME is operating in the range of applied load which forces the use of DSI. Therefore, to subjectively measure the effects of these impairments it is necessary to vary the applied load on the DCME up to and including the maximum design load. The clipping produced by the speech detector is affected by the type of signal being transmitted on the circuit under test. Therefore, only a realistic speech signal which also contains appropriate additive noise should be used on the circuit under test;

v) in most instances DCME is designed to be used in the network as a replacement for an existing facility. If the DCME introduces more delay than the facility replaced, then this additional delay will reduce the echo tolerance (grade of service) unless it is compensated for by the use of extra echo control measures. Based on Recommendation G.131, the magnitude of the reduction in the echo tolerance that will occur without extra echo control can be determined and hence a decision taken as to the need for additional echo control measures.

vi) The methodology should, ideally, yield results which can be used to produce new opinion models or modify existing models.

1.4 *Description of DCME*

Annex A contains a detailed description of the characteristics of DCME that can be evaluated with this methodology. This section contains a brief summary of these characteristics.

The test methodology applies to two types of DCME: one type which uses DSI only to obtain a DSI gain and a second type which uses a combination of LRE and DSI to obtain both a LRE gain and a DSI gain. The test methodology accounts for the operation of the speech detector, recognizing that speech clipping is an impairment that may occur even though the DCME is not overloaded.

The test methodology is applicable to DCME employing any one or a combination of three methods of overload control: 1) sample dropping or competitive clipping, 2) variable bit rate, and 3) queuing. The test plan also allows for testing of DCME having DLC capability.

The test methodology recognizes that many of the impairments produced by DCME occur only when a load is applied, and therefore provision is made to apply a controlled load to the DCME under test. The load is varied between zero and 100% of circuit capacity. Use of the packet mode in the DCME converts it into a packetized voice system, and this test methodology is applicable to these systems. At the present time only point-to-point (and possibly point-to-multipoint) DCME are covered by this methodology.

2 Source recordings

2.1 *Apparatus and environment*

The talker should be seated in a quiet room having a volume of between 40 and 120 cubic meters and a reverberation time of less than 500 ms (preferably in the range 200 to 300 ms). The room noise level must be below 30 dBA with no dominant peaks in the spectrum.

Speech should be recorded from an Intermediate Reference System (IRS), as specified in Recommendation P.48, or an equivalent circuit. The IRS is chosen as it is well documented and can be implemented by all laboratories. The IRS should be calibrated according to Recommendation P.64.

The recording equipment should be of high quality and of the type agreed to by the test. The equipment selected should be capable of providing at least a 40 dB signal-to-noise ratio. A suitable system might consist, for example, of a high-quality digital audio tape recording system.

All the source speech material should be recorded so that the active speech level, as measured according to Recommendation P.56, is approximately 23 dB below the peak overload level of the recording system. This will assure that the speech peaks will not overload the recording system.

2.2 *Speech material*

The speech material should consist of a sequence of simple, meaningful, short sentences, chosen at random because easy to understand (from current non-technical literature or newspapers, for example). Very short and very long sequences should be avoided, the aim being that each sequence when spoken should have a duration of at least 30 s and the duration of any two sequences should differ by no more than 5 s. Administrations can use one of two approaches:

i) to have as many different sequences as there are conditions (an example of suitable material from which sequences may be constructed is contained in Annex B), or

ii) to have a more limited number, e.g. 10 sequences per talker, where combinations of two sequences can be used (this is shown in detail in Annex C).

Because of the opinion scales to be used the first approach is recommended. Enough sequences should be available to cater for all the test conditions, plus a sufficient number for use in a practice session.

2.3 *Procedure*

At least three sentences should be used for each sequence. A silent period containing only circuit noise of approximately one second should procede the first sequence and the sequence should end with a similar silent period containing only the circuit noise. One of the inter-sentence pauses containing circuit noise should last one to two seconds. Otherwise, the talker should speak so that pauses occur naturally.

To facilitate the processing of the recorded speech through the DCME, i.e. to allow for the starting and stopping of the recorders between sequences and to allow time for adjusting the DCME for the next test condition, sequences should be separated by a 5 seconds gap on the tape. Therefore, the recorded source sequences will have the pattern on the tape shown in Figure 3/P.84.

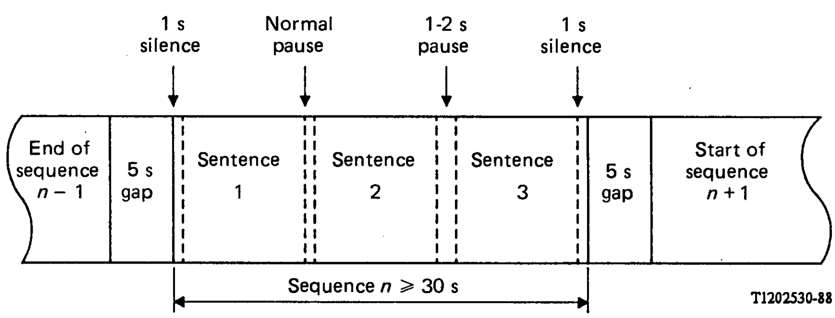


FIGURE 3/P.84

Sequences should be played back to listeners beginning with the one second silent period. After the sequence has ended, a 5 s period of complete silence should be provided to permit the listener to vote.

Talkers should pronounce the sequence of sentences fluently but not dramatically and have no speech deficiencies such as "stutter".

At least two male-female pairs of talkers shall be used, and more pairs are desirable if the test-time permits.

The method of presentation of the source sequences will be by randomization of talkers by blocks; as shown in the following example:

	Block 1	Block 2	Block 3	...	Block n
Talker	1 2 3 4	3 4 1 2	1 3 2 4		2 3 1 4

where talkers 1 and 2 are male and talkers 3 and 4 are female.

2.4 *Calibration signals and speech levels*

When the recordings have been made, the active speech level of each speech sequence (excluding the preceding and following silent periods) should be measured, preferably according to Recommendation P.56. If necessary, the speech should then be re-recorded onto the *right* channel of a second system with the necessary gain adjustments, so that all the sequences will be brought to the same speech level, namely 23 dB below the peak overload level of the recording system.

Thirty seconds of 1000 Hz tone should be inserted at the re-recording stage, at an r.m.s. level 17 dB above the active speech level, i.e. 6 dB below the peak overload level of the recording system: the peak level of this tone will be 3 dB higher still. This tone can then be used later to adjust the r.m.s. input speech level to be 20 dB below the overload point of the DCME (a peak/r.m.s. of tone of 3 dB with the speech level 17 dB below the r.m.s. tone level will give the 20 dB figure).

The *left* channel of the source recording should contain a 1000 Hz tone at a level 23 dB below the peak overload level and of 0.5 s duration, recorded about 0.5 s before the start and after the end of each sequence. These two signals may be used as checking and control signals in the processing of the source sequences through the DCME under test.

3 Simulating system load

3.1 *Requirements for a generic voice load simulator*

Digital Circuit Multiplication Equipment (DCME), by definition, is used to gain an advantage in the number of circuits multiplexed onto a digital transmission facility. With this advantage, however, comes potential degradation of transmission quality when carried loads exceed that for which the DCME was engineered. Thus, a rigorous performance evaluation of DCME includes studying the behaviour of the DCME under conditions of no load, engineered load, and overload. Because the transmission performance of DCME under load depends critically upon the load characteristics, it is necessary to use *known and controlled* simulated loads in order to properly assess DCME performance. This section describes the generic requirements for a voice load simulator for the purpose of facilitating DCME performance evaluations under conditions that are meaningful. Use of voice load simulators with the generic requirements described here will also enable the comparison of results from different studies of various DCME.

Note 1 – The load simulator specified here is to be used for the performance evaluation of DCME using Digital Speech Interpolation (DSI). This excludes Type A DCME, for which load is not an issue by virtue of the fixed time-slot assignment of the channels.

Note 2 – The load simulator specified here is an "external" simulator that produces simulated speech signals so as to exercise many circuits being multiplexed onto a digital transmission facility. Prototype DCME frequently use "internal" load simulation of "trunk needs service" requests that simulate the output of multiple speech detector circuits and thus compete for transmission capacity, even though no simulated signals are actually transmitted; only the "live" channel under test is actually transmitting. This type of simulator can be very useful in the lab, but is not treated here because certain assumptions would have to be made regarding the performance characteristics of the associated speech detector simulation.

3.1.1 *Parameters*

A generic Voice Load Simulator (VLS) for DCME performance evaluation has the following attributes (the parametric specification of which are detailed later in this section):

- talk-spurt characteristics,
- silence (gap) characteristics,
- background noise-fill for silent periods,
- spectral properties of the simulated speech,
- amplitude characteristics,
- physical interface, including idle-circuit specifications.

The above are a minimum set of parameters that may have to be expanded as required; for example, time variation of the number of simulated calls might have to be studied, at which time a pertinent specification would have to be added. Also, only simulated speech signals are discussed. It may be desirable to add simulated tones, signalling frequencies, and voiceband data of various types at a later date.

3.1.2 *Requirements*

3.1.2.1 *General*

These requirements apply to a generic VLS testing a DCME. Accordingly, the DCME must receive digital signals from the VLS that simulate multiple and independent sources of speech similar to that which is observed in telephone networks. To meet the "multiple and independent" condition, it will be assumed that the VLS output is to several T1 or CEPT interfaces.

Where possible, existing Recommendations have been used in deriving these requirements. The most notable exception are the requirements associated with speech activity and the underlying statistical distributions of talk-spurts and silent periods (gaps). For these, the current technical literature was surveyed; the results of [1] being both recent and based on conversational speech, are used here.

3.1.2.2 *Talk-spurt characteristics*

The probability density function (p.d.f.) of talk-spurt durations is modeled by two weighted geometric p.d.f.'s:

$$f_t(k) = C_1(1 - U_1)U_1^{k-1} + C_2(1 - U_2)U_2^{k-1}, \ k = 1, 2, 3, \ldots$$

where

$C_1 = 0.60278 \qquad U_1 = 0.92446$

$C_2 = 0.39817 \qquad U_2 = 0.98916.$

Every increment of the variable k is equal to 5 ms in time. The cumulative distribution function of talk-spurt durations is shown in Figure 4/P.84. The average talk-spurt duration is $\alpha = 227$ ms.

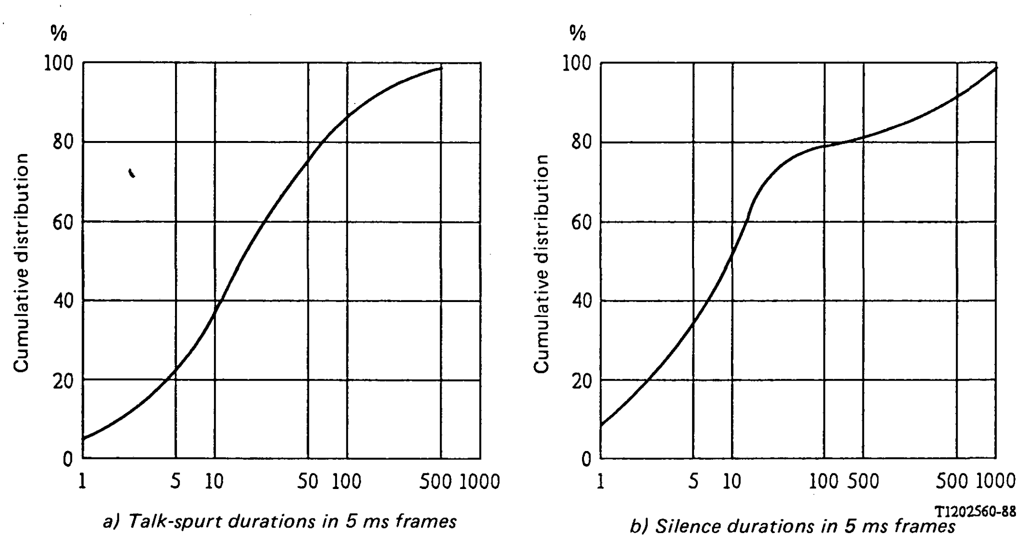


FIGURE 4/P.84

Probability distribution functions of talk-spurt and silence durations

3.1.2.3 *Silence (gap) characteristics*

The p.d.f. of silence durations is also modeled by two weighted geometric p.d.f.'s:

$$f_s(k) = D_1(1 - W_1)W_1^{k-1} + D_2(1 - W_2)W_2^{k-1}, \ k = 1, 2, 3, \ldots$$

where

$D_1 = 0.76693 \qquad W_1 = 0.89700.$

$D_2 = 0.23307 \qquad W_2 = 0.99791.$

The cumulative distribution function of silence (gap) durations is shown in Figure 4/P.84.

The average silence duration of $\beta = 596$ ms, combined with the 227 ms talk-spurt duration average, yields a long-term speech activity factor of 27.6 percent.

3.1.2.4 *Background noise-fill for silent periods*

Noise should be inserted into the silent periods (gaps) so that the performance of DSI in the presence of noise can be studied. It is desirable to have the noise level adjustable; a default value of −58 dbm0p is provisionally recommended.

3.1.2.5 *Properties of the simulated speech*

The artificial voice signal of Recommendation P.51 shall be used as a basis for simulating the characteristics of human speech. Supplement No. 7 to the Series P Recommendations describes a possible generation process of the artificial voice according to Recommendation P.51. This signal can then be switched on/off according to the talk-spurt and silence duration statistics described in §§ 3.1.2.3 and 3.1.2.4.

3.1.2.6 *Physical interface*

The load simulator should have T1 and/or CEPT outputs which have physical, electrical, coding, frame structure, alignment, and signalling characteristics as per Recommendations G.703, G.704, G.711 and G.732 (2048 kbit/s) or G.733 (1544 kbit/s).

3.2 *Determining load capacity of tested systems*

The average applied load equals the product of the number of circuits in use, N, and the average speech activity. The load capacity of the tested system equals the maximum load that the system is designed to handle, L_{max}. The load capacity can be determined by:

i) obtaining the manufacturer's specifications,

ii) calculation.

After the load capacity is determined, the partial loads at which the system will be tested can be determined. The partial loads are:

$$L_i = c_i \, L_{max}$$

where

c_i = 0.0, 0.50, 0.75 and 1.0.

3.3 *Controlling load applied to tested systems*

The load applied to the DCME can be changed by varying N and the activity factor. For these tests the speech activity factor will be assumed constant at 28%. Therefore, to obtain a partial load, L_i, it is necessary to calculate the number of active circuits which come closest to achieving this desired value.

For example, if $L_{max} = 48$ and if a partial load of $L_i = 0.50 \, L_{max}$ is desired and the speech activity factor of 28% is assumed, then the number of active circuits, N_{active}, is calculated thus:

$$N_{active} = c_i \frac{L_{max}}{\textit{(speech activity factor)}} = 0.5 \frac{48}{0.28} = 86 \text{ active circuits.}$$

In the test, 86 circuits would carry speech load and the remainder would be idled.

Note – The following items are for future study:

a) Should DCME loads include voiceband data as well as speech? The effect of voiceband data traffic on speech quality is an important issue in the evaluation of DCME performance. Data percentage is defined as follows:

$$P_{data} = \frac{\text{Number of input circuits active with data}}{\text{Total number of active circuits}} \times 100\%$$

b) Some Administrations report that speech activity on their real circuits averages about 36% when using a highly sensitive speech detector having a short hangover time of about 30 ms. Is it desirable to modify the speech load requirements given in § 3.1, and, if so, what values are recommended?

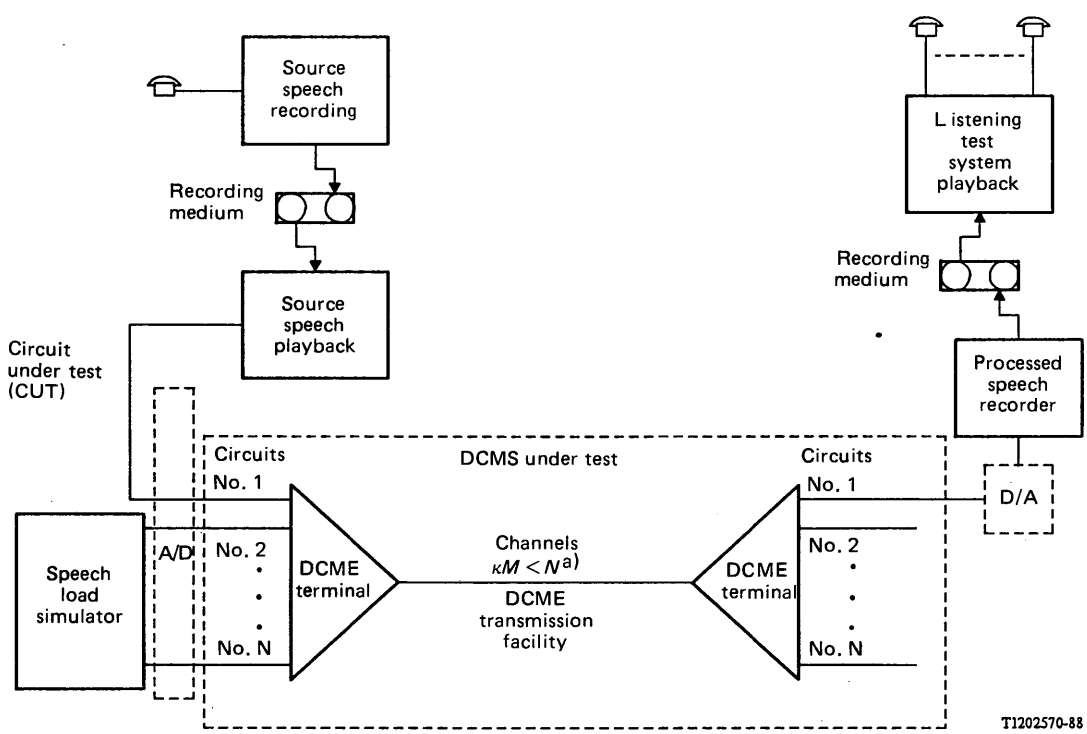
c) Fractional values of speech load are given in § 3.2. Some DCME may operate so as to display significant changes in performance at different fractional load points. Should the fractional load points be changed to accommodate this type of operation, and, if so, what changes are recommended?

4 Processing of the speech

The DCME testing laboratory will take the source recordings, replay them through the circuit under test of the agreed DCME (using the calibration tone to set the agreed input level), operating the DCME at the agreed load, and record the output from the circuit under test in a predetermined arrangement (explained in § 5). The recorded outputs will then be used to perform the listening test. The DCME being tested must be connected to the load simulator and to the recording and playback equipment as shown in Figure 5/P.84. It may be necessary to make provision for special A/D and D/A interfaces to permit the selected load simulator and recording equipment to be connected to the DCME.

All the processed outputs will be on the *left* channel of the recording medium. The corresponding original signal will be simultaneously recorded on the *right* channel. The 1 kHz tone will be available both in its original form (*right* channel) and as processed by passing through the DCME under test (*left* channel).

The 1 kHz tone on the source recording (see § 2) will be used to adjust the r.m.s. input speech level to be 20, 30 or 38 dB below the overload point of the DCME coder.



a) For an explanation of the factor κM, see § A.4.

FIGURE 5/P.84

Testing DCMS

5 Test design

Three separate tests are proposed to evaluate different aspects of DCME performance. The first verifies the effect of various loads on the performance. The second verifies the effect of errors in the DCME digital control channel. The third test calculates the effect that the DCME delay has on the echo tolerance. This last test will be done using Recommendation G.131 and does not involve subjective testing.

5.1 *Test No. 1: Effect of applied load*

This test may be conducted twice, once to obtain a quality rating and (optionally) a second time to obtain a listening effort rating. The parameters for testing are as follows:

a) *DCME test parameters:*

1. DCMEs under test: N
2. DCME loads: four values (0, 0.5, 0.75, 1.0) (see § 3.2)
3. speech activity factor: one value (28%)
4. active circuit speech characteristics: one value (see § 3.1)
5. circuit under test (CUT) idle circuit noise (ICN): two values[2)] (−77 and −45 dBm0p)
6. input speech level to CUT: three values (20, 30 and 38 dB below DCME coder overload)
7. output listening levels: at least three values (preferred and preferred ±10 dB)
8. talkers: four talkers, i.e. 2 male and 2 female.

b) *Reference parameters*

1. original source sequences: one value
2. MNRU: four values (5-35 dB in 10 dB steps)
3. SNR: three values (20, 30 and 40 dB)
4. reference connections (HRCs): approximately four different cases to be decided by test team
5. listening levels: three levels (see above)
6. talkers: four talkers, i.e. 2 male and 2 female.

For the stated set of parameters the number of test condition is:

$$4 \times 2 \times 3 \times 3 \times 4 \times N = 288 \times N \text{ DCME conditions}$$

plus

$$12 \times 3 \times 4 = 144 \text{ reference conditions.}$$

This totals (assuming $N = 1$ DCME):

$$432 \text{ test conditions} + 36 \text{ practice} = 468 \text{ conditions.}$$

The set of test conditions should be divided into about 13 segments (12 test + 1 practice) of 36 conditions with the conditions within each segment put into a random order. Table 1/P.84 lists the conditions in a basis non-randomized segment.

The basic balanced segment in Table 1/P.84 will be repeated for each of 4 talkers and 3 listening levels to create 12 test segments: A thru L. A practice segment P will also be created. The test segments A thru L plus P can then be ordered for playback in the listening test according to the procedure described in § 6.

Assuming each condition takes 35 s to present and obtain a vote, total test time is about 4.5 hours.

5.2 *Test No. 2: Effect of digital errors in the DCME control channel*

The preceding test was done assuming that the digital transmission facility is operated error-free. Under real conditions errors will occur and errors in the DCME control channel may cause momentary disruption of the voice circuits. To determine the effect of digital errors on performance, Test No. 1 should be repeated while random errors at a rate of 10^{-3} are injected into the control channel. For this test only one listening level (preferred) is necessary, so the total number of test conditions is $N \times 96$ plus 144 reference conditions. With $N = 1$, the test time is 2.3 hours.

---

2) Time permitting, use of a third noise level of −58 dBm0p is suggested. This will permit a better characterization of the effect different noise levels have on the DCME.

TABLE 1/P.84

Basic segment (assumes 1 DCME for testing)

Condition	Load	BCN (dBm0p)	Input [a] (dB)	Q (dB)	SNR (dB)	HRC
1	0.00	−77	20			
2	0.50	−77	20			
3	0.75	−77	20			
4	1.00	−77	20			
5	0.00	−45	20			
6	0.50	−45	20			
7	0.75	−45	20			
8	1.00	−45	20			
9	0.00	−77	30			
10	0.50	−77	30			
11	0.75	−77	30			
12	1.00	−77	30			
13	0.00	−45	30			
14	0.50	−45	30			
15	0.75	−45	30			
16	1.00	−45	30			
17	0.00	−77	38			
18	0.50	−77	38			
19	0.75	−77	38			
20	1.00	−77	38			
21	0.00	−45	38			
22	0.50	−45	38			
23	0.75	−45	38			
24	1.00	−45	38			
25			20			Original
26			20	5		
27			20	15		
28			20	25		
29			20	35		
30			20		20	
31			20		30	
32			20		40	
33			20			HRC1
34			20			HRC2
35			20			HRC3
36			20			HRC4

ICN idle circuit noise

[a] dB below DCME coder overload level.

5.3 *Test No. 3: Effect of delay*

In this test, using Recommendation G.131, the intent is to calculate the transmission delay through the DCME, then determine if the delay will require the use of additional echo control measures. The answer to this question requires that we define the connections that the DCME will be used to provide, then determine the echo tolerance of these connections assuming that conventional transmission facilities are used in place of the DCME, and then finally determine the reduction in the echo tolerance that will occur by inserting the DCME into the connections. If the reduction in tolerance falls below acceptable limits then additional echo control measures will be required if the DCME is used.

6 Listening test procedure

6.1 *Apparatus, calibration and environment*

The listening room should meet the same conditions as the recording room with the exception that the environmental noise should be set to 45 dBA (Hoth spectrum – Supplement No. 13, at the end of this fascicle.

The IRS receiving end (Recommendation P.48) or equivalent circuit will be used. The IRS should be calibrated according to Recommendation P.64.

The gain of the system should be set in such a way that the 1 kHz tone played back from the recordings produces a sound pressure of 7 dBPa when measured on the IEC 318 artificial ear (Recommendation P.51). Thus the speech level at that point will also be −10 dBPa (84 dB SPL) for undistorted speech which is close to the "preferred listening level".

6.2 *Instructions to subjects*

The instructions are given in Annex D. When the subjects have read these instructions, they should listen to the practice conditions and give their response to each sample. No suggestions should be made to them that the practice conditions exhaust the range of qualities that they can expect to hear. Questions about procedure or about the meaning of the instructions should be answered, but any technical questions must be met with the response, "We cannot tell you anything about that until the test is finished".

6.3 *Opinion scale*

The methods agreed to are both of the single stimulus type based on the mandatory "quality" scale and the optional "listening effort" scale.

6.3.1 *Opinions based on the "quality" scale*

The following five categories should be used for the quality test:

- Excellent
- Good
- Fair
- Poor
- Bad

or equivalent depending on language. (Supplement No. 2, at the end of this fascicle.

6.3.2 *Opinions based on the effort required to understand the meaning of sentences (listening effort scale)*

The following five categories should be used for the optional listening effort test:

- complete relaxation possible, no effort required;
- attention necessary, no appreciable effort required;
- moderate effort required;
- considerable effort required;
- no meaning understood with any feasible effort.

or equivalent according to language. (Supplement No. 2, at the end of this fascicle.)

Note 1 – It is expected that quality and listening effort scales are correlated. Therefore it is not generally required to use both scales. However, if, in a particular case, it is desirable to obtain ratings on both scales, the test should first be performed by using the listening effort scale and then duplicated using the quality scale. This order of presentation is particularly important if the same listeners and the same speech sources are used in both tests.

Note 2 – The rating scales associated with the categories defined in §§ 6.3.1 and 6.3.2 are assumed to be linear interval scales. It is recommended to bring this assumption to the attention of the subjects in the test instructions, either in words or by presenting numbers of numerical scales in the written instructions. Examples of how this can be done is given in Annex D. Alternatively, the scale can have more than 5 grades (e.g. 7 or 11 grades) with the same five verbal definitions at equal distances. An additional possibility is to define the end points of the scale separately (e.g. Ideal and Unusable). These defined end points then serve as anchoring points but are not supposed to be used for the rating. Examples of such alternative subjective scales are found in Annex E.

6.4 *Sequence of operations*

The 12 test plus 1 practice segments (A-L plus P) should be played back according to the augmented latin-squares:

Quality test	*Optional listening effort test*
P CABD ...	P ABDC ...
P DBAC ...	P DCAB ...
P ADCB ...	P BDCA ...
P BCDA ...	P CABD ...

In these squares, each row is used for each group of listeners, who may listen either simultaneously or separately. The segments are played back in the given order within each row. A pause will naturally occur between one segment and the next, while the right place on the recording medium is being found and possibly the calibration is checked; this pause will also be welcomed by the listeners.

6.5 *Listeners*

The listeners used in the tests should be drawn at random from the population of telephone service customers. About 40 but not less than 30 listeners should be solicited.

6.6 *Data collection*

Subject's responses may be collected by any convenient method: pencil and paper, press-buttons controlling lamps recorded by the operator, or automatic data-logging equipment, for example. But whatever method is used, care must be taken that subjects should not be able to observe other subjects' responses, nor should they be able to see the record of their own previous responses. Apart from the inevitable memory and practice effects, each response should be independent of every other.

7 **Statistical analysis and reporting of results**

After the test is finished and all subject responses are collected, the experimenter will assign numerical scores to the responses as follows:

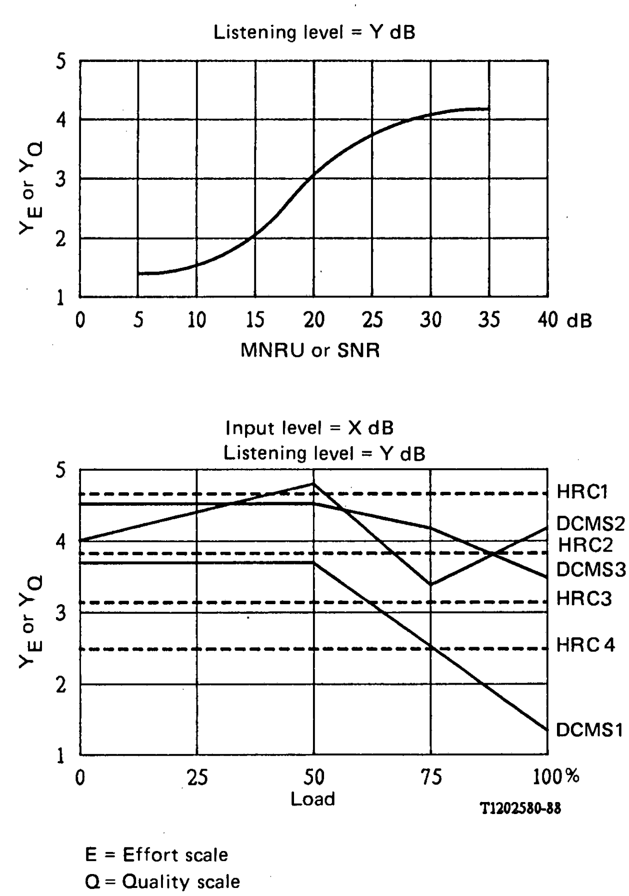
Response	*Score*
Excellent	5
Good	4
Fair	3
Poor	2
Bad	1
Complete relaxation possible, no effort required	5
Attention necessary, no appreciable effort required	4
Moderate effort required	3
Considerable effort required	2
No meaning understood with any feasible effort	1

The numerical mean score (over subjects) should be calculated for each condition, and these means listed (this is required so that effects due to male and female speech can be seen).

As a further aid to rapid review of results, graphs should be prepared according to the formats shown in Figure 6/P.84.

Note especially that the averaging of male and female results is here proposed purely to reduce the output to manageable proportions, and does not imply that this step would be warranted for the detailed study and interpretation of the results (unless the significance tests justify it).

Calculation of separate standard deviations for each condition is not recommended. Confidence limits should be evaluated and significance tests performed by conventional analysis-of-variance techniques.



Note 1 – There should be a set of 15 curves using the listening effort scale and another set of 15 curves for the quality scale.

Note 2 – Organizations are requested to use the axes as given above.

Note 3 – No significance should be attributed to the curves shown; they are for example only.

Note 4 – Generate one complete set of load plots for each of three input levels. Generate one complete set of all curves for each of three listening levels.

FIGURE 6/P.84

Format for presentation of results

ANNEX A

(to Recommendation P.84)

Description of digital circuit multiplication equipment

A.1 *Definition of DCME*

Digital circuit multiplication equipment (DCME) is defined in § 1.2.1. A working definition may be: any digital transmission method that derives more voicegrade circuits than is possible using equipment conforming to Recommendation G.711. For our purposes the term circuit may at times refer to a circuit between two switching points (trunk) or between the customers premises and a switching point (loop). At other times it may refer to an end-to-end digital connection. The circuit may also be physical or virtual. The term voicegrade means that the bandwidth of the circuit is nominally 3.1 kHz. We will attempt to avoid confusion by using suitable qualifiers, when necessary, to describe the kind of circuit we mean.

Based on the above definitions we conclude that there are three basic types of DCME. These are:

Type A – Uses only LRE (low rate encoding, < 64 kbit/s) to obtain a circuit multiplier larger than 1. Some LRE methods (e.g., 32 kbit/s ADPCM) are amenable to the subjective testing methods described in Recommendation P.70; other methods (e.g. 48 kbit/s vocoding) may require new subjective test methods.

Type B – Uses only digital speech interpolation (DSI) to obtain a circuit multiplier larger than 1. DSI is defined in § A.2. By definition the digital coding used in Type B DCME to derive a circuit, operates at 64 kbit/s and conforms to Recommendation G.711. Thus, the coding provides a circuit multiplier of unity. During periods of DCME overload any of several overload strategies may be used to resolve the contention for channels. The three basic overload strategies are defined in § A.5. For example, during momentary periods of overload the channel coding rate may be reduced to increase the channel capacity. However, this recoding action is attributed to the DSI and the circuit multiplier larger than 1 thus obtained is credited to the DSI.

Type C – Combination of Types A and B. This hybrid type employs LRE to obtain a circuit multiplier larger than 1, and then DSI to obtain an additional circuit multiplier larger than 1. For example, if the LRE comforms to Recommendation G.721 32 kbit/s ADPCM, then the coder has a circuit multiplier of $\kappa = 2$. The DSI may increase this multiplier by a further factor of 2 or 3, depending upon the DCME. The total multiplier, 4 to 6, is equal to the product of the LRE and DSI multipliers.

A.2 *Digital speech interpolation (DSI)*

Digital speech interpolation, is defined in § 1.2.8. A working definition of DSI may be: any method for assigning a voicegrade bearer channel on demand for the transmission of speech at the onset of the speech burst (talk-spurt). The bearer channel comes from a pool maintained by the DCME and the speech comes from an active circuit connected to the DCME. When the speech burst stops the channel is either:

i) relinquished and put back into the pool, or

ii) kept assigned to the circuit as long as the pool is not empty and the channel is not needed to service another circuit.

In the above context the term "bearer channel" refers to the transmission paths between the DCME terminals, which are used to carry the traffic on the circuits connected to the DCME. By definition, a bearer channel has the same bandwidth as a circuit, i.e. voicegrade. Bearer channels may be derived using time, space or even frequency or wavelength division multiplexing of the transmission medium used by the DCME. The transmission media may be copper wire, coaxial cable, radio path or fibre.

A.3 *Speech detection*

To perform DSI, the DCME must contain a speech detector. The speech detector monitors the circuits and determines when speech is present and when it is not. When speech is declared present the DCME attempts to assign an available bearer channel to the circuit. If no channel is available the DCME then invokes its overload strategy. When the speech burst ends the speech detector may provide some "hangover" to avoid tail-end clipping of the burst. Hangover extends the effective length of the burst.

"Fill-in" is another speech detector function sometimes employed to bridge or eliminate the silence gaps less than a certain length between speech bursts. Fill-in does not extend the length of individual bursts the way hangover does, but requires a processing delay equal to the maximum filled-in gap. Both hangover and fill-in increase the activity factor of the speech on the bearer channels.

To avoid front-end clipping of the speech burst, the speech detector sometimes employs delay of a few milliseconds to give it time to decide whether speech is present.

Clipping or mutilation of the speech burst (both front-end and possibly tail-end) may occur because the speech detector makes false or late decisions. The operation of the speech detector and thus the clipping performance of the DCME is a function of many factors characterizing the signal on the circuits, such as the signal level, signal-to-noise ratio, and echo path loss.

A.4 *Definition of load*

The frequency of DCME overloading is a function of the load on the system. The system load consists of the speech bursts generated on the incoming circuits plus DCME generated overhead traffic. Since the speech burst activity on the circuits varies from moment to moment, the load also has short-term variations.

In defining load we must distinguish between the applied load and the offered load. The applied load is the speech bursts entering the DCME on the circuits in use. Thus, applied load is a function of the number of circuits in use and the speech activity on the circuits. The offered load consists of the applied load plus any additional load generated by the DCME. The offered load is the load presented to the DCME channels. It should be evident that the offered load is usually larger than the applied load, because:

i) the speech detector increases the activity factor, since it adds fill-in or hangover to speech bursts;

ii) "overhead" information may have to be transmitted on the channels along with the speech samples.

While the load varies continuously, subject to the statistics of the speech and the circuit activity, if we assume that the number of circuits in use, N, is a constant over some period of time in which we are observing the operation of the DCME, then the average applied and offered loads becomes useful concepts. Formulas for the average loads are defined in § 1.2.19. While these formulas are somewhat simplistic and do not capture the information concerning the variance of the load about the average, they do allow useful insight into the operation of the DCME.

The load carrying capacity of the DCME channels is also an important consideration. The load carrying capacity is defined as the maximum offered speech plus "overhead" load that the DCME channels can carry. If the offered load is less than the load carrying capacity of the channels, then all the offered load is carried by the DCME. However, if the offered load exceeds the capacity of the channels, then depending upon the overload strategy of the DCME, (see § A.5) some of the offered load will be lost through sample dropping, or variable bit rate coding will be used to momentarily increase the load carrying of the channels so that they can accommodate the extra load. Thus, overloading is defined to occur when the offered load exceeds the carrying capacity of the DCME channels.

In a sample dropping system the load capacity is fixed and is simply κM, where M is the number of 64 kbit/s equivalent channels provided and κ is the LRE factor which accounts for the difference in bit rates between the circuits (always 64 kbit/s) and the channels. If 32 kbit/s LRE is used on the channels, for example, then $\kappa = 2$. If LRE is not used then $\kappa = 1$. If variable bit rate (VBR) coding is used then the load capacity of the DCME is not fixed, and overloading may be avoided by temporarily creating extra bearer channels. If the coding rate drops from 32 to 16 kbit/s, for example, then during the period VBR is active $\kappa = 4$.

In these examples the number of channels available to carry speech is assumed to be constant. However, in DCME that carries voiceband data and other tones on the circuits, DSI cannot be used on these signals. The result is that these continuous signals capture channels for full-time use, reducing the pool of channels available for carrying speech.

By using the average load equations and the concept of load capacity, we can illustrate in Figure A-1/P.84 the load curves for a sample dropping type C DCME. The slope of the offered load curves depends upon the speech activity factor. $\alpha/(\alpha + \beta)$, and the speech detector "stretch" factor, k. Load curves for three different activity factors are shown. If the number of circuits in use, N, is less than $N_{\min} = \kappa M - G = 43$ then the DSI will never activate, even if the momentary speech activity factor goes to unity on all active circuits. Since the DCME-carried load cannot exceed $\kappa M = 48$, as the average offered load, L_o, gets closer and closer to the maximum capacity, the frequency of overloading (sample dropping) will increase as the moment-to-moment fluctuations in the speech activities push the offered load above the limit.

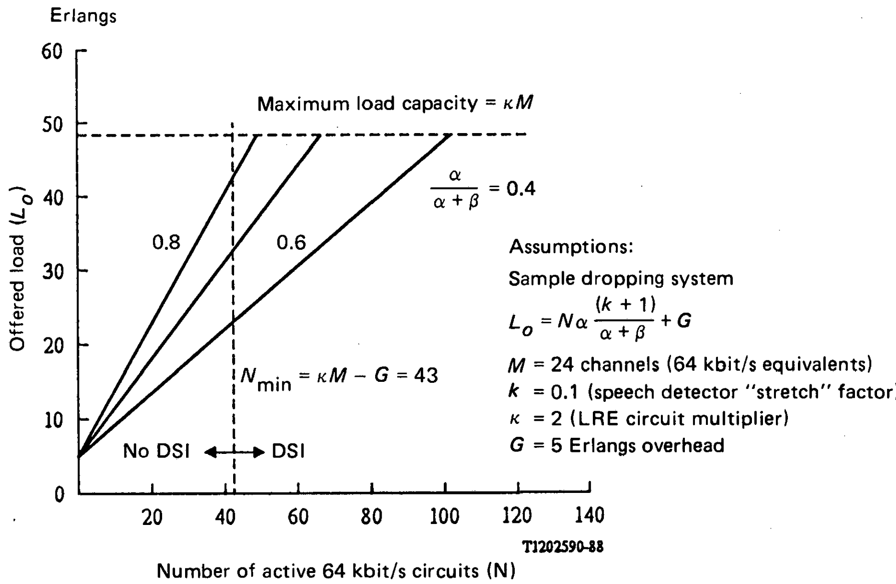


FIGURE A-1/P.84

Example of load curves for type-C, sample dropping, DCME

Figure A-2/P.84 illustrates the load curves for a variable bit rate type C system which recodes at 16 kbit/s during overload. In this example, when the offered load exceeds $\kappa M = 48$ the coding rate is dropped from 32 to 16 kbit/s on the bearer channels. The capacity is thus increased to $\kappa M = 96$. The extra capacity absorbs the momentary overload and prevents sample dropping (freezeout) from occurring. If the offered load exceeds 96 then sample dropping will have to occur, because further VBR (e.g. down to 8 kbit/s) is not provided for in this example.

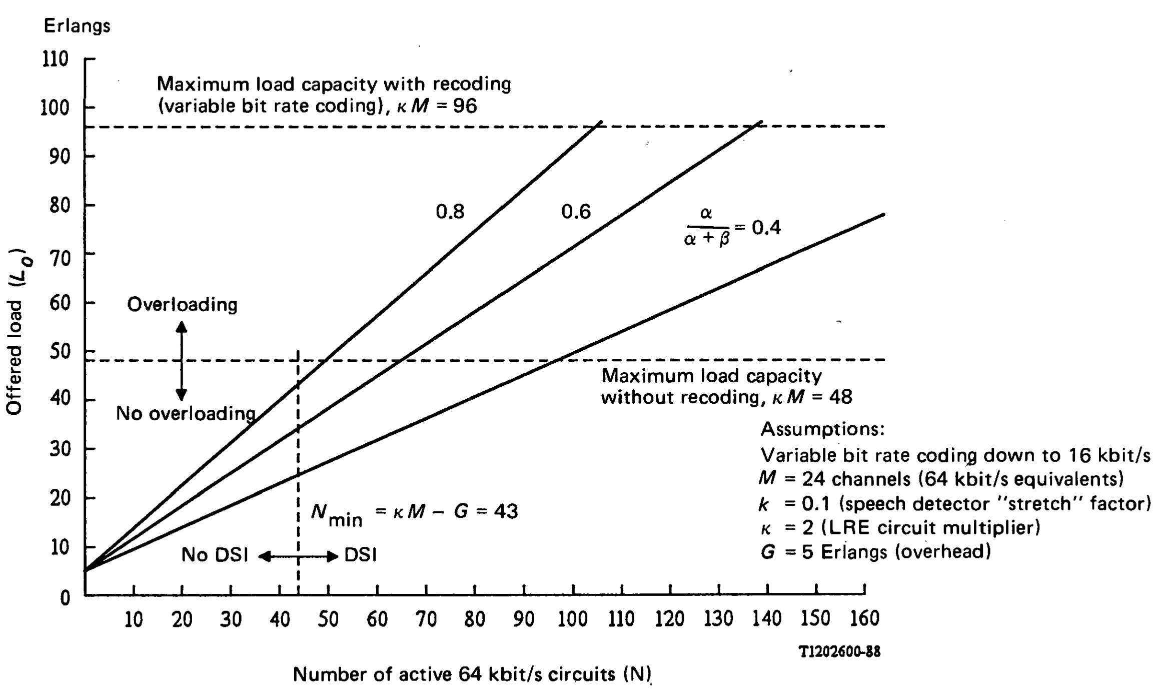


FIGURE A-2/P.84

Example of load curves for type-C, variable rate coding, DCME

Thus, in summary, as long as $N \leqslant N_{\min}$ the DCME will not need to use the DSI function, because all circuits will have access to a bearer channel. Overload will not occur until the offered load exceeds the load carrying capacity. In overload, the DCME will start dropping samples or will queue the samples, in which case κ will not change, or the DCME will decrease the coding rate, in which case κ will increase, thus momentarily increasing the capacity of the DCME.

A.5 *Overload strategies*

When a number of active circuits connected to the DCME exceeds the number of available channels, the DCME will experience momentary overloads; an increase in speech bursts will sometimes require more channels than are available. When this happens the DCME must invoke its "overload strategy". The strategy is designed to deal with the issue of how best to share the channel pool. A number of basic strategies are possible:

Type 1 – Competitive clipping or speech sample dropping. In this strategy, defined in § 1.2.14, samples are dropped from the front end of the speech burst that unsuccessfully bids for a channel. Sample dropping continues until a channel is available or the burst ends normally. Perceptually, the effects of front-end sample dropping and front-end clipping, the latter caused by the speech detector, should be the same, even though they have different causes. Theoretically, however, they are not entirely the same, because front-end clipping is more likely to affect low-level parts of the signal, whereas freezeout affects all levels with equal probability.

Type 2 – Variable bit rate coding. This strategy, defined in § 1.2.15, employs embedded speech coding algorithms or other means to effectively multiply the number of bearer channels momentarily available to the circuits to carry the offered load. Since a lowering of the bit rate will have the effect of increasing the quantization noise produced by the coders, the perceptual effect of variable rate coding will be momentary increases in quantizing noise, i.e. reductions in Q (for a discussion of Q, see Recommendation P.81, § 2).

Type 3 – Queueing. This strategy, defined in § 1.2.16, employs buffers (memories) for the speech burst samples to occupy while waiting for a channel. The perceptual effect of pure queueing, without buffer overflow, is a time shift of the speech bursts. No samples are lost, and there is no increase in noise. The impairment introduced can be called "silence duration modulation". From the listener's point of view a given speech burst when queued will begin somewhat later in time relative to its predecessor burst than it would have without queueing. Also the succeeding burst may be perceived as beginning somewhat sooner. Since the buffers must, of necessity, be finite this strategy cannot be employed alone, but it must be coupled with either sample dropping or variable rate coding. Thus, a queueing system can have speech mutilation or recoding noise as well as time shifting.

Type 4 – Dynamic load control. An overload control strategy, defined in § 1.2.17, in which the DCME signals to the associated switch that the traffic load which the switch is generating, or is predicted to generate, cannot be transmitted satisfactorily by the DCME, and the switch should reduce its demand on the DCME by a holding signal sent to the circuits when they become idle.

A.6 *Silence reconstruction methods*

Since the DCME does not transmit silences between speech bursts at the receiving end, the silences must be artificially recreated. Several different methods for doing this are possible. The simplest is to insert a white noise at a fixed level in the receiver during silences. Careful selection of the level is necessary to avoid noise contrast, that is, an apparent and annoying contrast between the noise in the silences and the background noise during speech bursts. Other methods are possible which attempt to adapt the noise level automatically to the circuit conditions; these methods require careful filtering and estimation of source noise power.

A.7 *Circuit versus packet mode*

Internally the DCME may employ a circuit or a packet mode for the transmission of speech bursts. In the circuit mode, bearer channels are derived by providing suitable time slots on the transmission facility interconnecting the DCME terminal equipment. In the packet mode, the speech burst samples are put into one or more packets of fixed or variable length. The packets are addressed to the destination circuit and transmitted over the transmission facility one at a time. Thus, in the circuit mode the transmission facility can be thought of as carrying a number of channels multiplexed together, while in the packet mode the facility is thought of as a single high speed channel which transmits packets one at a time.

In the packet mode, performance of the system depends on how the packets are serviced. Two servicing methods are:

a) All packets from all circuits enter a first-in first-out (FIFO) queue and are serviced by the high speed channel one at a time. Each packet is treated independently. Each packet experiences a variable delay in arriving at the receiving end that is a function of the fill of the FIFO queue. If packets arrive too late, after a given reconstruction delay, they will be lost or discarded by the receiver. This is called packet dropping and it is a function of the system load. Packet dropping can cause speech mutilation at any point in the burst. It gives rise to "mid-burst" sample dropping. Packets can also be dropped in the FIFO queue if it experiences overflow. The fill of the queue is monitored and the overload strategy is invoked when necessary to prevent excessive packet dropping.

b) Once a circuit has seized the high speed channel for transmission of a packet all the packets on the circuit for that burst are transmitted before the high speed line is free to transmit another circuit's packets. Thus the circuit is "cut-through" during the burst. Cut-through operation avoids mid-burst speech sample loss. However, since only one circuit at a time can use the high speed channel, other circuits with packets to transmit must await their turn. The packets must be queued while they await the channel. Load-dependent queueing delays must be equalized at the receiving end. This is usually done by employing some form of time stamp on the packet. The possibility always exists that packet queues will overflow before the packets can be transmitted. When this happens the overload strategy is invoked to prevent excessive packet dropping.

Packet mode introduces more delay than a non-packet mode DCME. The extra delay has three components. The first is the packetization time. Packetization time is a function of packet length and circuit coding rate. The second is the reconstruction delay. Reconstruction delay is chosen to minimize the probability of packet loss. The third is packet queueing delay. All three components may be load-dependent and thus variable.

In summary, use of packet mode rather than circuit mode may introduce these additional performance-affecting aspects:

i) mid-burst sample dropping,

ii) additional delay equal to the sum of the packetization and reconstruction delays,

iii) packet queueing delay.

A.8 *Packet reconstruction*

In a packet mode, system loss of a packet presents the receiver with a dilemma, namely, what to use in place of the speech samples carried in the lost packet. Several methods are employed and they have different performance consequences. One method is to insert noise samples in place of the lost speech samples. Another method repeats samples in a previous packet to replace the lost samples. Other methods are also employed.

A.9 *Circuit versus network systems*

With the above definitions in mind there appears to be yet another way to classify DCME. We can talk about DCME using non-switched channels and DCME using switched channels. The first type, non-switched channels, is called a circuit-based DCME. The second type, using switched channels, is called a connection-based DCME.

A circuit-based system would be used to provide circuits, either trunks or loops. All switching is done outside the DCME. The connection-based system incorporates circuit- or packet-switching and thus is more properly thought of as a network solution rather than a circuit solution.

The testing of a connection-based DCME is likely to be more complicated than is the testing of a circuit-based DCME. One reason is that the size of a connection-based system may make it difficult to test in a laboratory. Another reason is that loading such a system with a controlled load is difficult.

ANNEX B

(to Recommendation P.84)

Speech material used to construct speech sequences
(The following narratives are examples used by Bell Communications Research)

ORWELL

George Orwell began his classic novel 1984 with, "It was a bright cold day in April," but he gave no further hint as to what the weather might be during the fateful year. From the succession of untoward weather events that marked 1983, many have come to believe that the world's weather has undergone an unprecedented change for the worse and that we might be headed for a series of natural disasters this year to match the demise of free democratic thought and speech described in Orwell's book.

Since we do not have the ability to predict what individual weather events might occur during 1984, let us turn the calendar back a hundred years and see what happened throughout the country in 1884. The year opened with the arrival of arctic air from northern Canada which drove the thermometer down to $-40°$ F at Rockford, Illinois, and to $-25°$ F at Indianapolis, Indiana, both records that still stand. Sub-zero temperatures penetrated into the South, and a hard freeze hit citrus groves in Florida.

In early February, heavy rains falling on a deep snow cover caused the Ohio River to flood. Crests were of record height from Cincinnati to the river's mouth at Cairo, Illinois.

Late February brought an outbreak of tornados in the South and the Ohio Valley, where some sixty individual funnels descended. More than 420 were killed, and more than 1000 injured. Nothing approached this visitation in severity or extent until the tornado outbreak in April in Durango, Colorado, for seventy-six days ending April 16.

In May, out-of-season rainstorms in the deserts of the Southwest caused widespread floods. Rail traffic from Salt Lake City to the south was interrupted for three weeks, and the Rio Grande River flooding at El Paso, Texas, caused $1 million in damage.

Heavy frosts occurred in late May, when the thermometer dropped to 22° F in Massachusetts, and snow fell in Vermont on Memorial Day.

California got more heavy rain in June; Los Angeles had 1.39 inches and San Francisco 2.57 inches, both all-time June records. And as a result of rain in Wisconsin the flooding Chippewa River did more than $1.5 million in damages and left 2,000 homeless at Eau Claire.

The great Oregon snow blockade followed 34 inches of snowfall at Portland in the middle of December. Rail communication was cut off from the east and south for many days, and mail from California had to come by ocean steamer.

If you think the weather that made so many headlines in 1983 was unprecedented, hark back to 1884. We do not know whether El Nino was active then or whether some other atmospheric or oceanic force was the culprit. All we can do now is wait and see what 1984 brings.

FOG

One of winter's most spectacular sights is a smokelike fog that rises from openings in the arctic ice fields and occasionally appears above the open waters of unfrozen lakes and harbors in our temperate zone. Various names for the phenomenon are "frost smoke", "sea smoke", "steam fog", "warm water fog", and "water smoke". The fog is caused by the passage of a stream of arctic or polar air with a temperature near zero Fahrenheit over unfrozen water. Within the lower forty-eight states, it occurs principally over unfrozen areas of the Great Lakes and over harbor waters of the north Atlantic coast.

"Sea smoke" occurs because the vapor pressure at the surface of the water is greater than that in the air above. Water vapor evaporates into the air faster than the air can accommodate it. The excess moisture condenses and forms a layer of fog, like steam or smoke rising off the water. Usually a clear space exists between the water's surface and the bottom of the fog, and its upper limit is generally 10 to 25 feet. If an atmospheric inversion develops near the water's surface, the fog may be confined there and becomes thick, resulting in a hazard to navigation.

If the air temperature is severely cold, $-20°$ F or below, the rising moisture may form ice crystals in the layer of air just above the water. This is called "frost smoke", and it makes a beautiful sight, especially when sunlight glitters on the thin ice needles.

"Steam fog" can occur over lakes and streams in the autumn following a clear, still night during which the air has cooled. The differences in vapor pressures cause the warm water to steam into the cold air, and whole valleys and basins can be covered with a thin layer of fog while the hillside remains clear.

ANNEX C

(to Recommendation P.84)

Instructions on the use of a limited number of sentences

(Contribution by the Swedish Telecommunication Administration)

If N sentences per talker are used there will be $N(N-1)$ possible sentence combinations per talker. The first 16 results are tabulated below:

N	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$N(N-1)$	2	6	12	20	30	42	56	72	90	110	132	156	182	210	240	272

Either of two reasons for wanting to limit the number of sentences can be put forth:

- the wish to save time by not having to author lists of more than 2×85 sentence combinations per talker. Separate recording of all the combinations is of course still needed unless sophisticated editing equipment for digital types is at hand, or
- the need to organize the test in a way that fulfills the requirements for an analysis of variance.

Depending on which of the motives above is invoked, different methods can be adopted. These are:

1) All possible $N(N-1)$ sentence combinations per talker are recorded.
   a) The same N sentences are used for all 4 talkers. The same sentence pair should then not be used for the same test conditions from talker to talker, in order to avoid possible systematic interaction between test conditions and phonetic content, or
   b) Four different sets of N sentences ($N1$, $N2$, $N3$ and $N4$) are authored. Then no precautions corresponding to a) are needed. However, interaction will still be possible and uncontrolled.
2) To allow for an analysis of variance, subjects must judge the same speech material for all test conditions and all talkers. The number of sentences will then be limited to $M \times 2$ where M is the number of pairs that will be used in the test. If $M = 1$ the test may appear too tedious for the subjects and the phonetic coverage may be insufficient. If an analysis of variance is to be justified, and the test is still to be practically possible, an expansion of the number of presentations is therefore recommended. $M = 2$ or 3 should be enough. This will lengthen the test time for each subject, but experience shows that tests of 2.5 hours per subject are quite possible. Adjustments for such an expansion must then be made when deciding the presentation order.

ANNEX D

(to Recommendation P.84)

Instructions to subjects

D.1 *Quality scale – DCME test*

In this test we are evaluating systems that might be used for telecommunications service between separate places.

You are going to hear a number of samples of speech reproduced in the earpiece of the handset. Each sample will consist of a 30 to 35 seconds long sequence of three or more sentences.

Please listen to the complete sequence, then indicate your opinion of the overall sound quality. If you hear any noises or other interference in the pauses before, between or following the sentences you should include the effect of this interference in your judgement of the overall quality.

For indicating your opinion you are requested to use the following 5-point rating scale:

Score	*Quality opinion*
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad or Unsatisfactory

After listening to a sample sequence, either (1) please write down on your response sheet a score, or (2) please press the appropriate button which on this rating scale represents your opinion of the sound quality of the sample just heard.

After you have given your opinion there will be a short pause before the next sample begins.

For practice, you will first hear "*n*" samples and give an opinion on each; then there will be a break to make sure that everything is clear.

From then on you will have a break after every "*k*" samples. There will be a total of "*t*" samples in the test. The test will last a total of about "*time*" hours.

D.2 *Listening effort scale – DCME test*

In this test we are evaluating systems that might be used for telecommunications service between separate places.

You are going to hear a number of samples of speech reproduced in the earpiece of the handset. Each sample will consist of a 30 to 35 seconds long sequence of three or more sentences.

Please listen to the complete sequence, then indicate your opinion of the effort required to understand the meaning of the sentences.

For indicating your opinion you are requested to use the following 5-point rating scale:

Score	*Listening effort opinion*
5	Complete relaxation possible, no effort required
4	Attention necessary, no appreciable effort required
3	Moderate effort required
2	Considerable effort required
1	No meaning understood with any feasible effort

After listening to a sample sequence, either (1) please write down on your response sheet a score, or (2) please press the appropriate button which on this rating scale represents your opinion of the effort required to understand the meaning of the sample just heard.

After you have given your opinion there will be a short pause before the next sample begins.

For practice, you will first hear "*n*" samples and give an opinion on each; then there will be a break to make sure that everything is clear.

From then on you will have a break after every "*k*" samples. There will be a total of "*t*" samples in the test. The test will last a total of about "*time*" hours.

ANNEX E

(to Recommendation P.84)

Examples of other subjective scales

E.1 *Eleven-grade quality scale*

10		
9	Excellent	The number 10 denotes a reproduction that is perfectly faithful to the ideal. No improvement is possible.
8		
7	Good	
6		
5	Fair	
4		
3	Poor	The number 0 denotes a reproduction that has no similarity to the ideal. A worse reproduction cannot be imagined.
2		
1	Bad	
0		

(See IEC Report 268-13, Annex A.)

E.2 *Seven point quality scale*

Score	*Quality description*
6	Ideal circuit
5	Excellent circuit. Possible to relax completely during call, very agreeable
4	Good circuit. Necessary to pay attention, but not necessary to make a special effort. Agreeable circuit
3	Fair circuit. A moderate, but not too great, effort is necessary. Not a very agreeable circuit
2	Poor circuit. Listening is possible, but somewhat difficult. Listening disagreeable
1	Bad circuit. Can be used only with great difficulty. Listening very disagreeable
0	Very bad circuit. Practically unusable

(See CCIR Report 751, Volume VIII.3, 1986.)

E.3 *Five-grade impairment scale*

5 Imperceptible.

4 Perceptible, but not annoying.

3 Slightly annoying.

2 Annoying.

1 Very annoying.

(See Supplement No. 14, Annex B.)

Reference

[1] LEE and UN: A study of ON-OFF characteristics of conversational speech, *IEEE Trans. Comm.*, Vol. COM-34, No. 6, June 1986.

PAGE INTENTIONALLY LEFT BLANK

PAGE LAISSEE EN BLANC INTENTIONNELLEMENT

PART II

SUPPLEMENTS TO SERIES P RECOMMENDATIONS

PAGE INTENTIONALLY LEFT BLANK

PAGE LAISSEE EN BLANC INTENTIONNELLEMENT

Supplement No. 1

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

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

Supplement No. 2

METHODS USED FOR ASSESSING TELEPHONY TRANSMISSION PERFORMANCE

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

(Quoted in Recommendation P.80)

(Contribution from British Telecom)

1 Introduction

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

2 List of methods

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

b) articulation (AEN) ratings;

c) listening opinion tests;

d) conversation opinion tests;

e) quantal-response detectability tests.

3 Brief descriptions and references to more complete descriptions

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

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

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

3.3.1 *Method of conducting listening opinion tests*

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

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

– Loudness preference scale:

Opinion scale No. 4A

A Much louder than preferred.

B Louder than preferred.

C Preferred.

D Quieter than preferred.

E Much quieter than preferred.

– Listening effort scale:

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

A Complete relaxation possible; no effort required.

B Attention necessary; no appreciable effort required.

C Moderate effort required.

D Considerable effort required.

E No meaning understood with any feasible effort.

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

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

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

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

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

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

3.4.1 *Method for conducting conversation tests*

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

3.4.1.1 *Experimental design*

The most suitable designs are of the $n \times n$ graeco-latin square type, where each of n pairs of subjects carries out one conversation on each of n circuit conditions. Precision is very low if n is less than 8; at the other extreme it is not practical to expect subjects to attend on more than four occasions, or to carry out more than four conversations per visit. Moreover, the total number of conversations, $n \times n$, increases much more rapidly than n. For this reason, n is normally limited to the range 8 to 15 inclusive: graeco-latin squares (with symbols from two alphabets) exist for all these numbers. In such a design, the convention is that rows denote pairs of subjects; columns denote the order of administering the experiment; symbols of the first alphabet denote circuit conditions (distinguished not only according to properties of the connections by themselves, but also according to room noise levels and any other "treatment" factors); symbols of the second alphabet denote sets of pictures used as the topic of conversation. No further orthogonal factors can be incorporated at all where $n = 10$, 14 or 15, nor by any simple method when $n = 12$; but where $n = 8$, 9, 11 or 13, it is possible to construct hyper-graeco-latin squares with symbols from $(n - 3)$ additional alphabets, which may be used to govern further orthogonal factors (such as selection of carbon microphones, choice of calling party, or choice of crosstalk recording), for each conversation. When the square is not hyper-graeco-latin these factors must be allocated by some simple balanced rotation scheme, but this may give rise to biases that cannot be eliminated from the results. For this reason the recommended value of n is now 13 rather than 12 as previously.

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

3.4.1.2 *Choice of circuit conditions*

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

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

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

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

3.4.1.3 *Eligibility of subjects*

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

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

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

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

3.4.1.4 *Environment*

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

3.4.1.5 *Methods of establishing the connection*

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

3.4.1.6 *Conversation task*

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

The subject is instructed to consider these pictures for display in a public place, and, before each conversation, to arrange the cards of a particular set in his personal order of preference for this purpose; the other subject does the same with his copy of the same set. When contact is established via the test circuit, the subjects have to negotiate an agreed order of preference and write this down at the end of the conversation. The duration of each conversation is thus determined by the subjects themselves. Occasionally a conversation may be very long because both subjects are intensely interested in the pictures, or – as happens in less than 1% of cases – very short because both have independently chosen the same order of preference and have little to discuss, but even in these cases it is highly desirable to allow the subjects to decide for themselves how long to converse. After the end of the conversation they express independent opinions of the connection by marking a form provided: one version of this form is reproduced in Annex A.

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

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

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

The experimenter prepares schedules, based on the experimental design, showing in what order conditions must be administered to each pair of subjects, with which picture sets, which party initiates the call in each case, and any other necessary details. Space is left for filling in information that becomes available as the experiment proceeds: consecutive conversation number, duration of conversation, identity of tape reel used for recording, comments about faults or unusual events, and so on. Opinion forms (Annex A) are also prepared for each conversation. However, in order to avoid duplicating or altering too many entries, some items are not filled in until they are certain: for example, the actual names of the subjects are liable to change until they actually arrive for their first visit.

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

3.4.1.8 *Procedure*

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

At the beginning of each conversation, the subjects take out the specified picture set from a box on the desk, arrange the pictures in order of preference, and fill in the appropriate part of the opinion form. When both subjects have done this, the experimenter gives one of them the signal to initiate the call. The subjects are then completely free to determine the course of the conversation, except that they must not discuss their opinions of the connection. When they have written down their agreed order of preference for the pictures, terminated the conversation, and recorded their scores (Excellent, Good, Fair, Poor or Bad) and their answer to the "Difficulty" question (Yes or No), the experimenter contacts each in turn by telephone to ask what answer he has given to the "Difficulty" question; if the answer is "Yes", the experimenter asks the subject to explain briefly (in his own words) the nature of the difficulty. The reply is noted, but neither the subject nor the experimenter is expected to attempt precise formulations: it is essential not to prompt the subjects, and in any case the classification of difficulty has been found far less useful than the undifferentiated percentage "Difficulty" itself.

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

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

3.4.1.9 *Treatment of results*

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

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

3.5 *Quantal-response detectability tests*

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

Opinion scale 6A

A Objectionable
B Detectable
C Not detectable

where B is understood to mean "Detectable but not objectionable".

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

It is sometimes permissible to regard these votes as opinion scores, with values 2, 1, 0 respectively, and treat them in the same sort of way as one would treat listening or conversation opinion scores. But this is often unsatisfactory because the decisions on such a scale as 6A are not really equivalents of responses on a continuous scale – as votes on such scales as 4A may be legitimately taken to be – but effectively embody two distinct dichotomies (for example detectable/not detectable and objectionable/not objectionable), which though not independent may nevertheless call different psychological processes into action: in other words, Objectionability or Intelligibility differs in kind, not merely in degree, from Detectability, and often has a different standard deviation. For this reason a more profitable method of analysis is to express the probability of response according to each dichotomy separately, as a function of some objective variable, by fitting probit or logit equations, and then using the quantiles or other parameters as a basis of comparison between circuit conditions, in a manner analogous to that used in applying articulations scores.

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

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

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

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

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

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

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

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

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

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

5 General comments on subjective methods used in the laboratory

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

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

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

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

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

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

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

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

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

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

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

6 Objective methods

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

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

ANNEX A

(to Supplement No. 2)

Opinion form 12A

Test ____________________

Name ____________________ Cabinet ____________________ No. ____________________

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

Your Order of Preference	1st	2nd	3rd	4th	5th	6th

2 If you receive a green GO AHEAD signal, then call your partner on ______________. Otherwise wait for your partner to call you.

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

	1st	2nd	3rd	4th	5th	6th
Partner's Order of Preference						

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

	1st	2nd	3rd	4th	5th	6th
Agreed Order of Preference						

Then replace your handset.

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

Excellent	Good	Fair	Poor	Bad

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

YES	
NO	

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

FOR R13.4 USE														

ANNEX B

(to Supplement No. 2)

(Standard letter sent to subjects)

Name ____________________

Group ____________________

R13.4 SUBJECTIVE TEST No. ____________________

Thank you for agreeing to take part in this experiment.

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

Time	Day	Date

On arrival, ask for __________________, quoting the above subjective test number. You will be reminded by telephone shortly before each visit is due. You may book your time to project __________________. If you cannot keep an appointment, or if you need further information, contact __________________ on Ipswich 64 _________________.

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

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

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

Thank you once again for your co-operation.

____

____________________ (date)

References

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

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

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

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

[5] *Ibid.*, § 4.5.1.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Bibliography

BRAUN (K.): Die Bezugsdämpfung und ihre Berechnung aus der Restdämpfungskurve (Frequenzkurve) eines Übertragungssystems; *T.F.T.*, Vol. 28, pp. 311-318, August 1939.

BRAUN (K.): Theoretische und experimentelle Untersuchung der Bezugsdämpfung und der Lautstärke; *T.F.T.*, Vol. 29, pp. 31-37, No. 2, 1940.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Supplement No. 3

MODELS FOR PREDICTING TRANSMISSION QUALITY FROM OBJECTIVE MEASUREMENTS

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

1 **Transmission rating models** *(Geneva, 1980; modified at Malaga-Torremolinos, 1984) (Quoted in § 3 of Recommendation P.11)* (Contribution by the Bell Communications Research, Inc. [1)])

1.1 *Introduction*

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

The models for circuit noise overall loudness rating (OLR) and talker echo are based on several conversational tests conducted at Bell Laboratories in the period from 1965 to 1972 to evaluate the subjective assessment of transmission quality as a function of circuit noise, overall loudness rating, talker echo path loss and talker echo path delay [1]. These tests involved several hundred subjects and several thousand test calls, Several tests were conducted on normal business calls. Others were conducted in the laboratory. All of the tests employed a 5-category rating scale: excellent, good, fair, poor and unsatisfactory.

---

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

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

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

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

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

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

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

All of the tests were conducted with Western Electric 500-type telephone sets or equivalent. The procedures used in the analysis of the subjective tests results and the derivation of the transmission rating scale are outlined in Reference [1]. Although the procedures are somewhat complex for manual calculation, they are easily handled on a digital computer and have been found to provide a convenient and useful representation for a large variety of test data.

The models incorporate the concept of a transmission rating scale. An important reason for the introduction of this scale was the recognition that subjective test results can be affected by various factors such as the subject group, the type of test, and the range of conditions which are included in the test. These factors have been found to cause changes in both the mean opinion score of a given condition and in the standard deviation. Thus, there are difficulties in trying to establish a unique relationship between a given transmission condition and subjective opinion in terms of mean opinion score or percent of ratings which are good or excellent. The introduction of a transmission rating scale tends to reduce this difficulty by separating the relationship between transmission characteristics and opinion ratings into two parts. The first part, the transmission rating as a function of the transmission characteristic, is anchored at two points and tends to be much less dependent on individual tests. The second part, the relationship between the transmission rating and subjective opinion ratings, can then be displayed for each individual test.

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

TABLE 1-1

Overall loudness rating (dB)	Circuit noise (dBmp) [a]	Transmission rating
16	−61	80
31	−76	40

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

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

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

1.2 *Transmission rating models*

1.2.1 *Overall loudness rating and circuit noise*

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

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

where

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

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

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

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

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

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

1.2.2 *Circuit noise equivalent of the room noise*

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

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

where

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

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

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

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

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

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

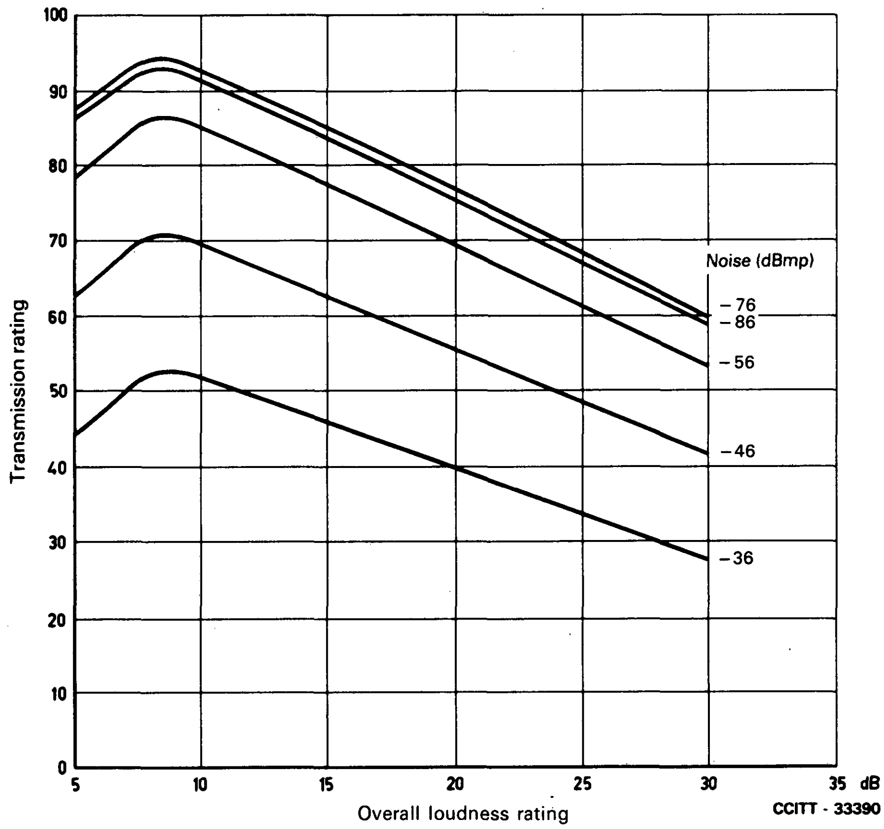


FIGURE 1-1

Transmission rating for OLR and circuit noise

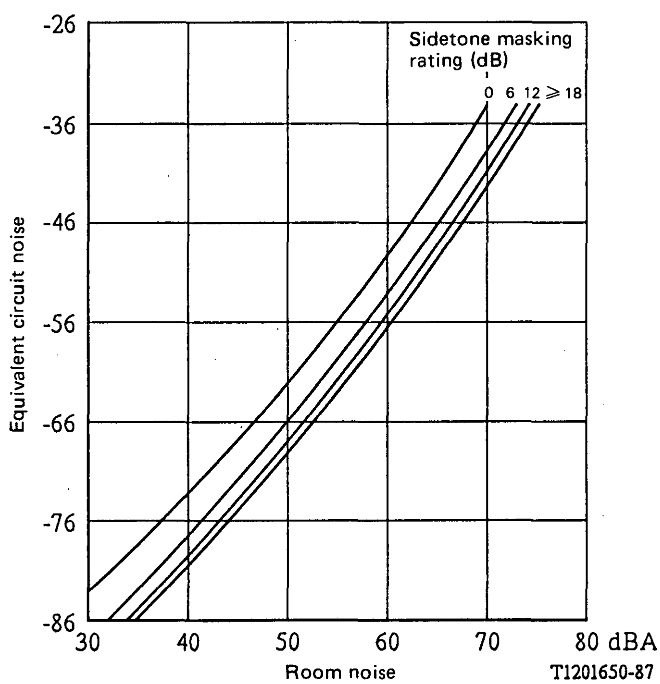


FIGURE 1-2

Equivalent circuit noise for room noise

1.2.3 *Circuit noise equivalent of quantizing noise*

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

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

where

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

and

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

SNR can be approximated by

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

from which

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

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

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

from which

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

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

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

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

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

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

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

where

L is the line bit rate in kbit/s.

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

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

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

1.2.4 *Bandwidth and attenuation distortion*

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

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

where

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

with

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

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

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

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

and

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

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

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

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

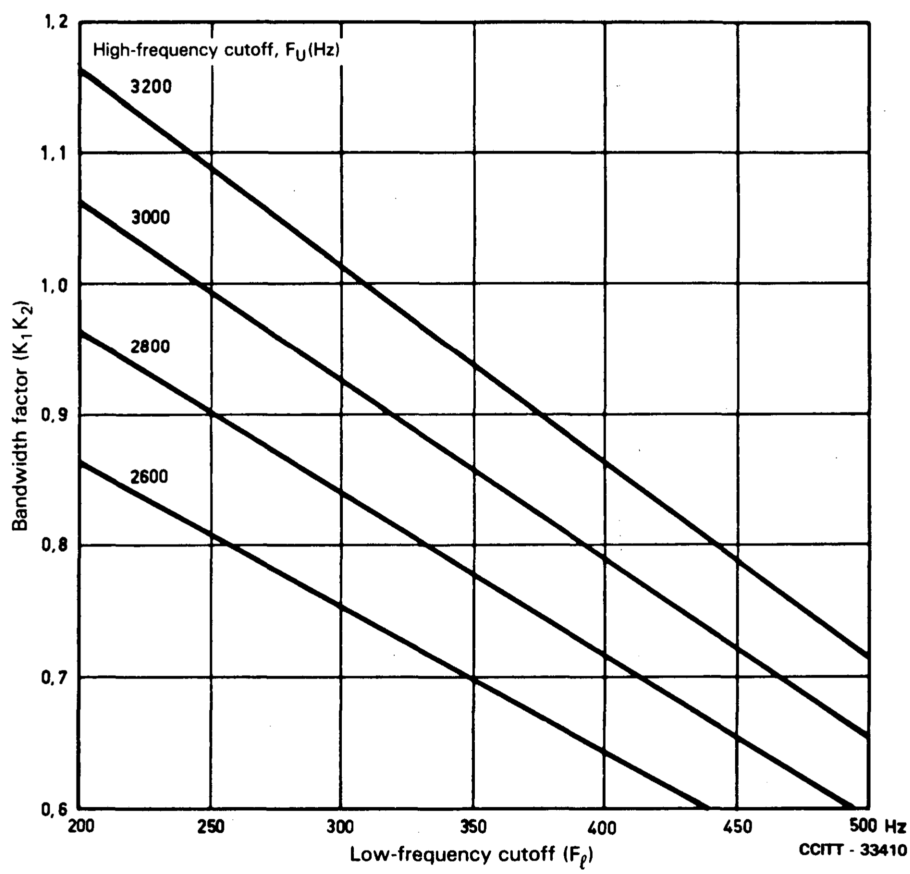


FIGURE 1-3

Bandwidth model factor

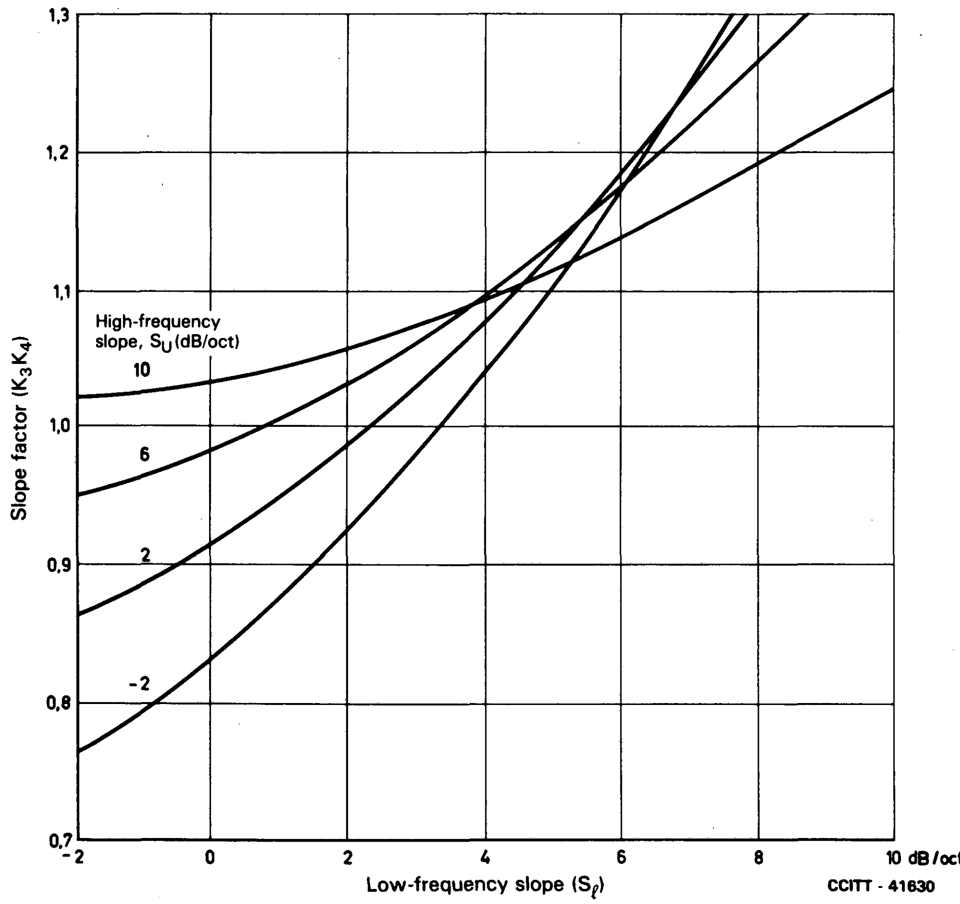


FIGURE 1-4
Attenuation-distortion model slope factors

1.2.5 *Listener echo*

The transmission rating model for listener echo is

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

where

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

$$WEPL = -20 \log_{10} \frac{1}{3200} \int_{200}^{3400} 10^{-\frac{EPL(f)}{20}}\, \mathrm{d}f \tag{1-21}$$

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

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

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

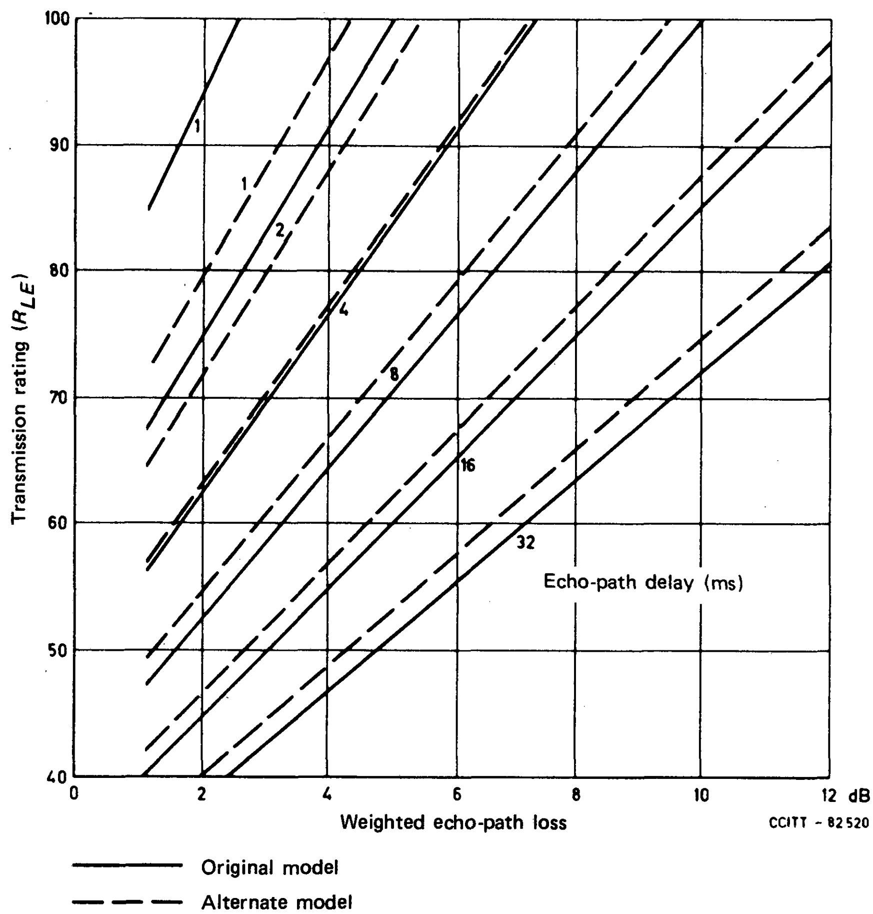


FIGURE 1-5

Transmission rating for listener echo

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

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

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

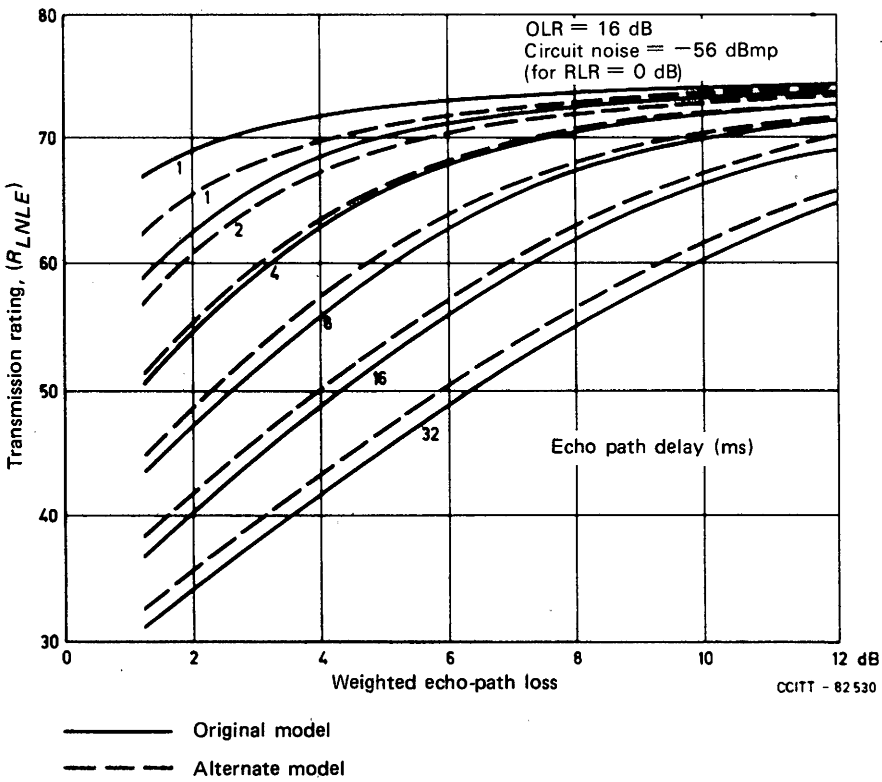


FIGURE 1-6

Transmission rating for OLR, circuit noise and listener echo

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

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

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

$$WEPL = SM + SF$$

where

SM is the singing margin,

SF is the shape factor

and then defined

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

Hence, like *WEPL*,

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

Also,

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

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

1.2.6 *Talker echo*

The transmission rating model for talker echo is

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

where

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

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

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

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

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

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

1.2.7 *Sidetone*

The transmission rating model for OLR, total effective noise and talker echo can be modified to include the effects of sidetone. The transmission rating, R_{LN-ST}, for OLR, total effective noise and sidetone is

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

and for talker echo and sidetone is

$$R_{E-ST} = R_E + 2.6(12 - SL) - 1.5(4.5 - SR)^2 + 3.38. \qquad (1\text{-}26)$$

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

$$K_{ST} = 1.021 - 0.002(SL - 15)^2 + 0.001(SR - 2)^2(SL - 15). \qquad (1\text{-}27)$$

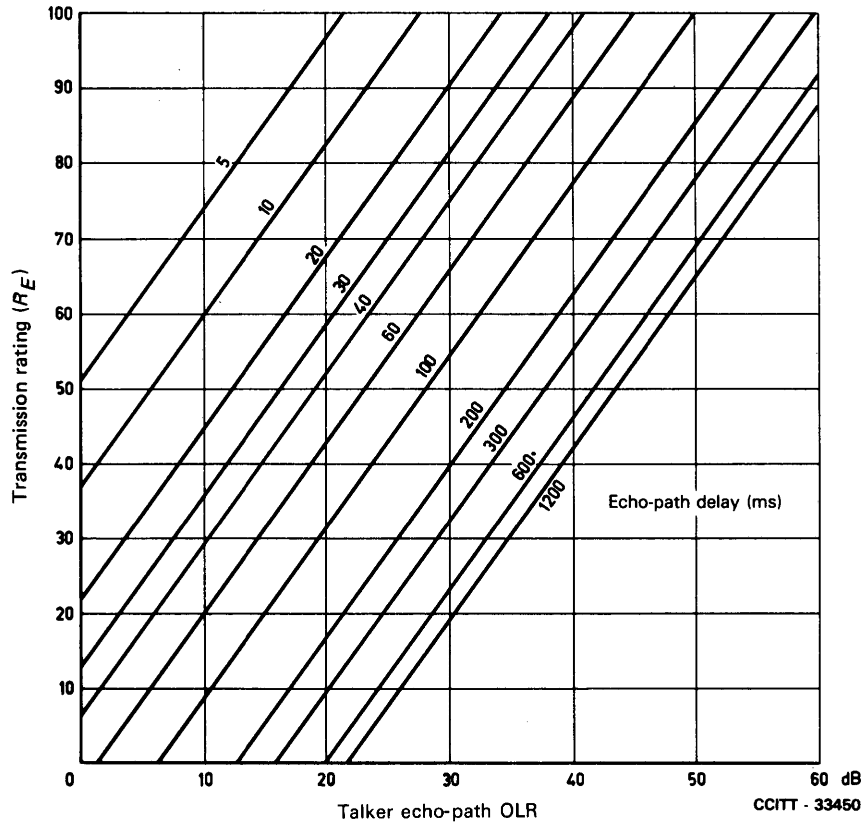


FIGURE 1-7

Transmission rating for talker echo

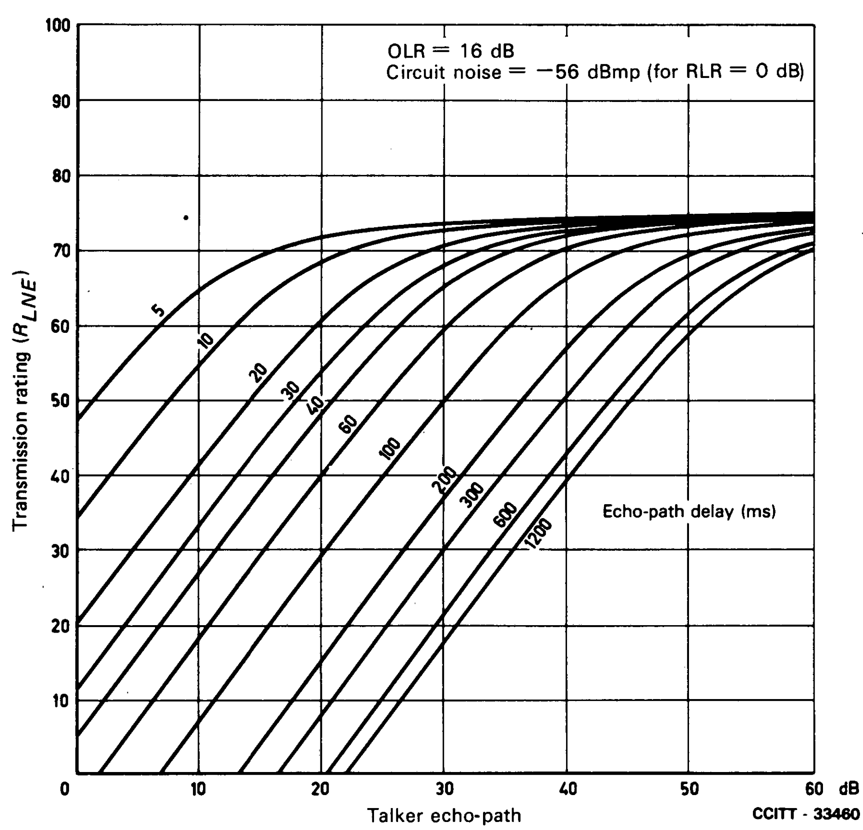


FIGURE 1-8

Transmission rating for OLR, circuit noise and talker echo

SL is the sidetone masking rating (in dB), *SR* is the sidetone response (in dB/octave) below 1 kHz. (The sidetone response above 1 kHz is 1.5 times greater.[2])

Figure 1-9 shows curves obtained by determining R_{LN-ST} and R_{E-ST}, then substituting these values for R_{LN} and R_E respectively in equation (1-24).

1.3 *Subjective opinion models*

Subjective opinion in terms of the proportion of ratings in each of the five categories (E, G, F, P, U) for a condition having a given transmission rating has been found to depend on various factors such as the subject group, the range of conditions presented in a test, the year in which the test was conducted, and whether the test was conducted on conversations in a laboratory environment or on normal telephone calls. The proportion of comments Good plus Excellent (G + E) or Poor plus Unsatisfactory (P + U) can be computed from the following equations:

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

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

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

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

Data Base[3]	*A*	*B*
1965 Murray Hill SIBYL Test	(R-64.07)/17.57	(R-51.87)/17.57
CCITT Conversation Tests	(R-62)/15	(R-43)/15
Long Toll Interviews	(R-51.5)/15.71	(R-40.98)/15.71

2 **Prediction of transmission qualities from objective measurements** *(Geneva, 1980; modified in Malaga-Torremolinos, 1984) (Quoted in Recommendation P.11)* (Contribution from British Telecom)[4]

Summary

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

---

[2] *Sidetone Response:*

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

[3] The three data bases reflect different relationships between the transmission rating scale and opinion ratings as determined in different tests as indicated below:

1965 Murray Hill SIBYL Test – Opinions on actual intra-building business calls.

CCITT Conversation Tests – Composite model of opinion in laboratory conversation tests reported to the CCITT in the 1973-1976 Study Period (see [3]).

Long Toll Interviews – Opinions expressed by North American Telephone customers when interviewed following a call on a long toll connection.

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

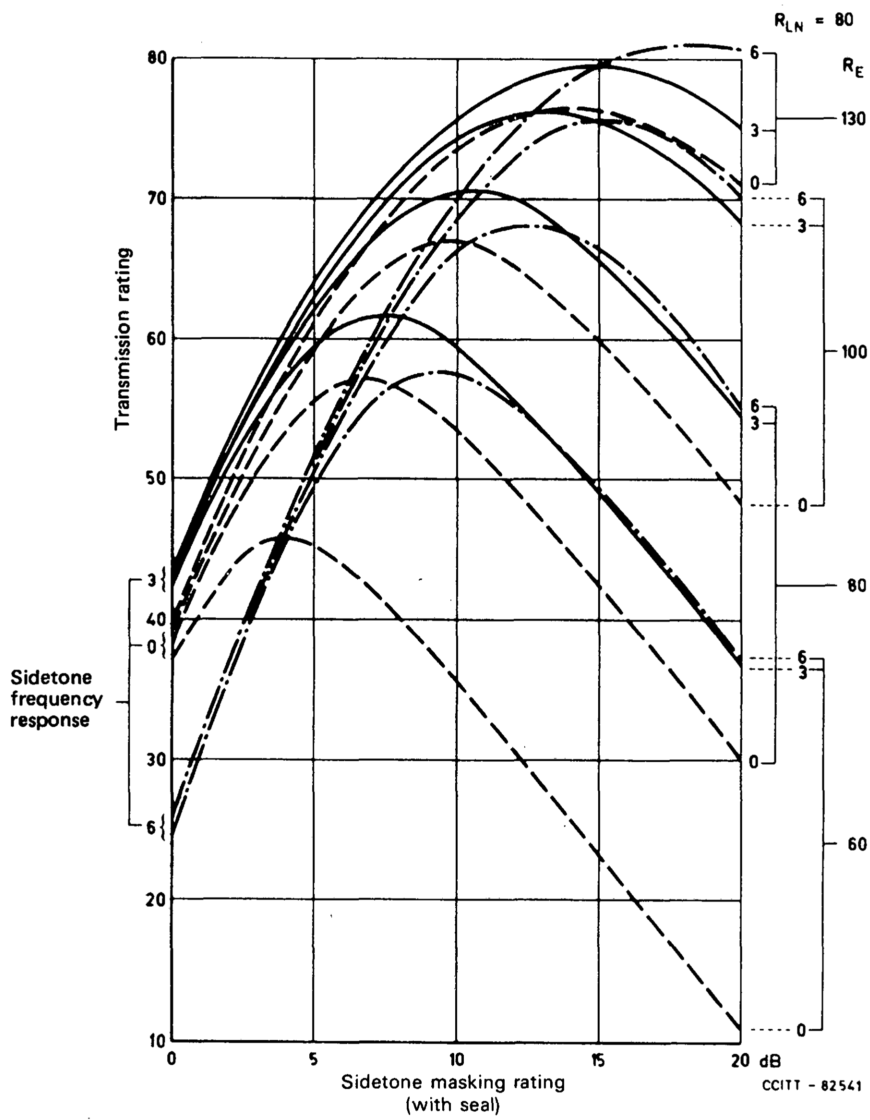


FIGURE 1-9

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

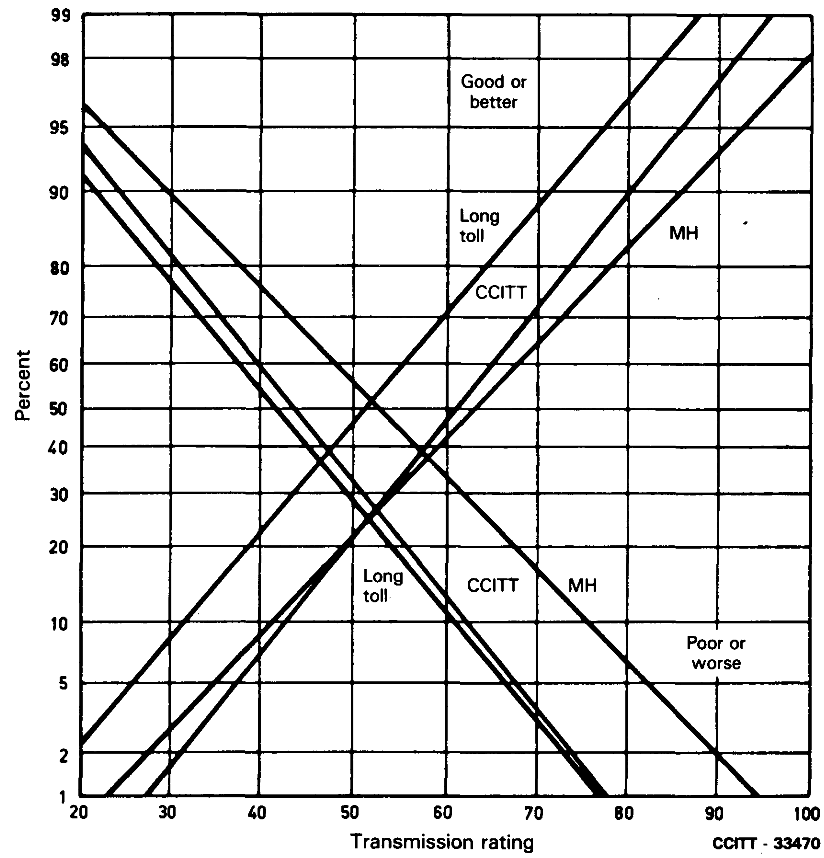


FIGURE 1-10

Comparison of opinion ratings as a function of transmission rating

2.1 *Types of model*

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

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

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

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

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

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

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

a) minor improvements to the subjective model;

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

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

2.3 *Situation to be represented*

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

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

Effort required to understand the meanings of sentences

A Complete relaxation possible; no effort required

B Attention necessary; no appreciable effort required

C Moderate effort required

D Considerable effort required

E No meaning understood with any feasible effort.

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

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

2.4 *Outline of the model*

The model requires the following inputs:

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

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

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

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

2.5 *Calculation of loudness and loudness ratings*

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

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

2.6 *Calculation of listening effort score*

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

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

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

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

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

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

where

- B'_i is the frequency weighting for the ith band, (shown diagrammatically in Figure 2-1),
- Z_i is the mean Z in the i th band,
- P is the appropriate growth function (illustrated in Figure 2-2),
- A is a multiplier depending on the received speech level, with the value 1 for a small range of levels around the optimum but decreasing rapidly outside this range (see Figure 2-3 where the zero abscissa now corresponds to OLR = 8 dB (Recommendation P.XXE [20]) instead of 4 dB as previously),
- D is a multiplier depending on the received noise level (ICN-RLR) with a value decreasing slowly from 1 at negligible noise levels towards 0 at very high levels (see Figure 2-4).

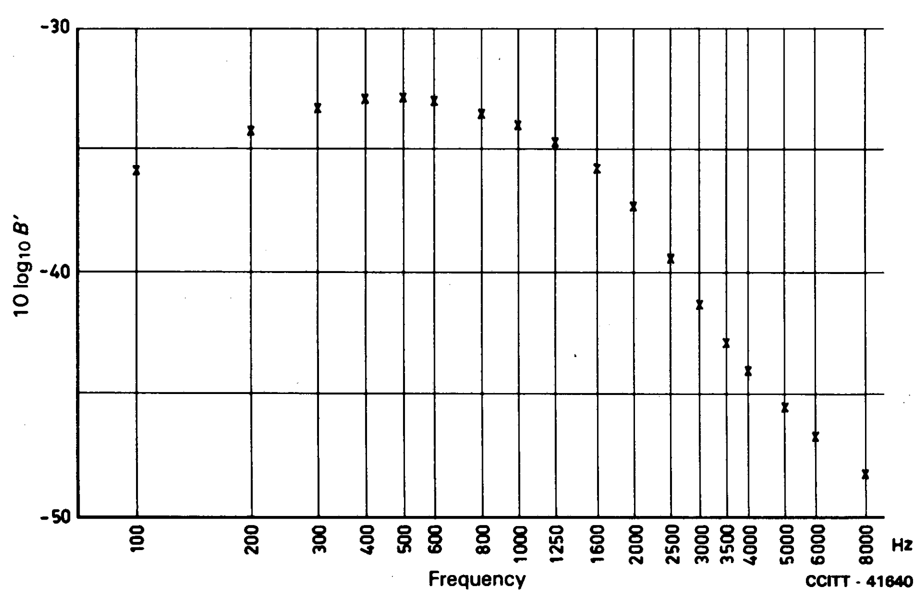


FIGURE 2-1

Frequency-weighting factor B' for Listening Opinion Index

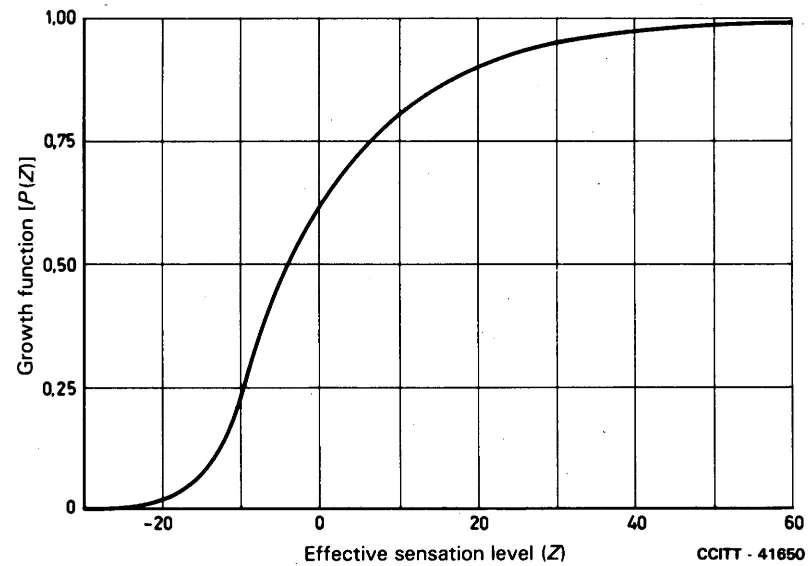


FIGURE 2-2

Growth function $P(Z)$

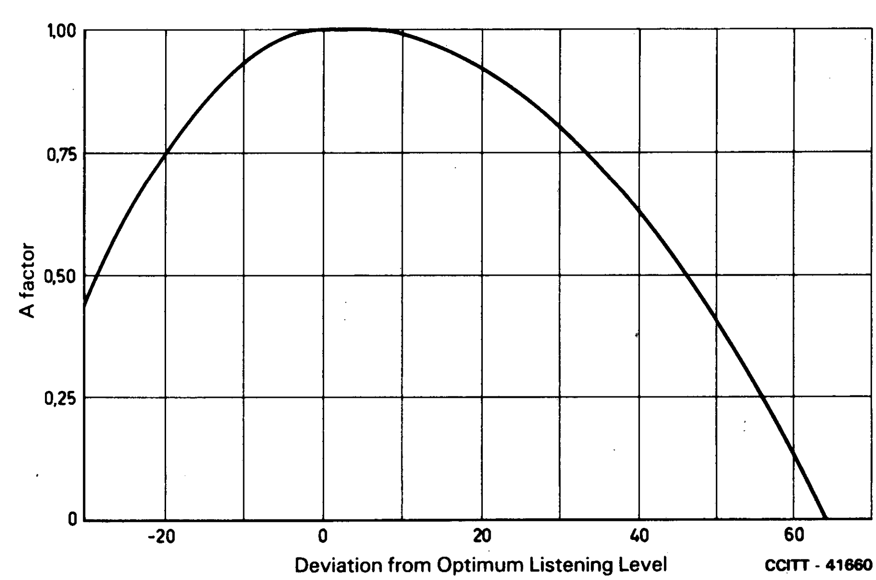


FIGURE 2-3

Effect of listening level on Listening Opinion Index

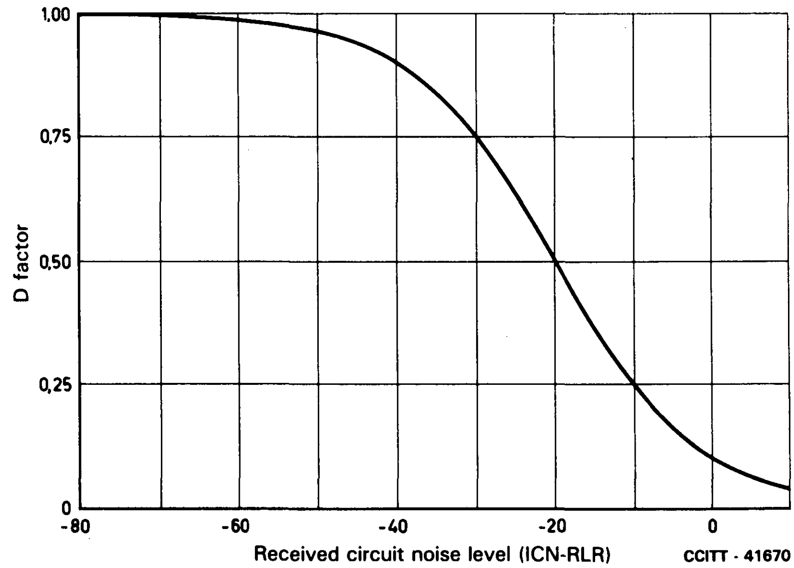


FIGURE 2-4

Effect of received noise level on Listening Opinion Index

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

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

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

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

2.7 *Calculation of Conversation Opinion Score*

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

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

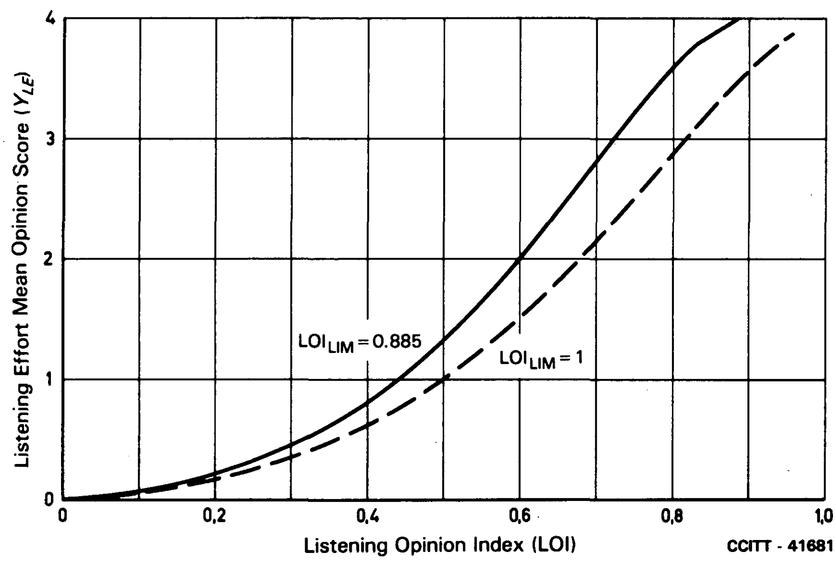


FIGURE 2-5
Listening Opinion Score as a function of Listening Opinion Index

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

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

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

Y'_C is an intermediate quantity explained below.

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

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

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

where

K = 1 if STMR < 13,

K = 0 otherwise.

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

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

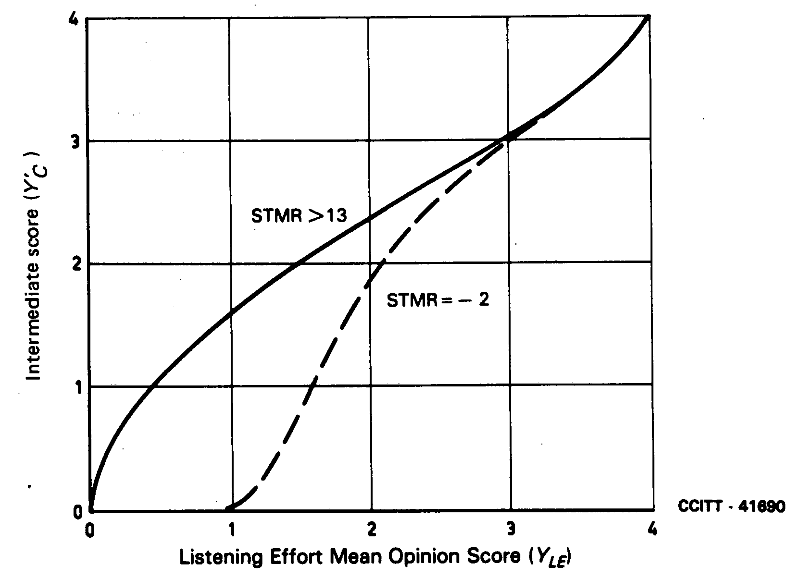


FIGURE 2-6

Intermediate score as a function of Listening Opinion Score

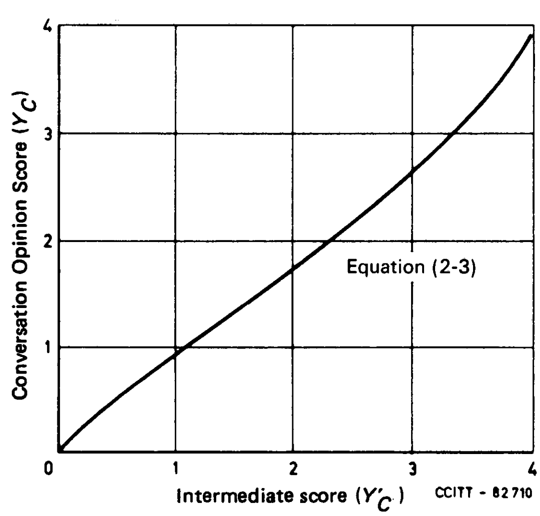


FIGURE 2-7

Conversation opinion score as a function of intermediate score

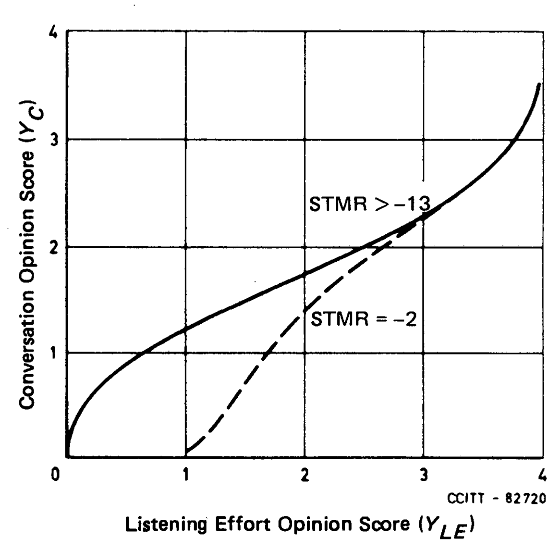


FIGURE 2-8

Conversation Opinion Score as a function of Listening Opinion Score

2.8 *Evaluation of other subjective measures of performance*

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

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

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

where

$D \times 100 = \%D.$

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

2.9 *Correspondence between calculated and observed values*

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

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

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

TABLE 2-1

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

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

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

2.10 *Incorporating miscellaneous degradations*

2.10.1 *PCM quantizing distortion*

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

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

Thus for a connection involving PCM links, one must include an evaluation of equivalent noise level in the iterative process that determines V_C: each successive approximation to V_C leads to a new value for Q, hence to a new value for G, and hence to a new contribution to the circuit noise to be taken into account in calculating the new value of Y_{LE}. In practice these modifications have negligible effect unless the speech level at the input to the PCM system falls below about -25 dBV, or the circuit noise at the same point is very high, or the speech input level is so high (say > -5 dBV) that appreciable peak limiting occurs.

2.10.2 *Syllabic companding*

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

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

TABLE 2-2

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

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

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

$$\begin{aligned}\text{Total improvement} &= 1/3\ (\text{improvement in presence of speech}) + \\ &\quad + 2/3\ (\text{improvement in absence of speech}) \\ &= -S/6 - 2N/3.\end{aligned}$$

Hence

$$\begin{aligned}\text{equivalent noise level} &= N - \text{improvement} \\ &= N + S/6 + 2N/3 = S/6 + 5N/3.\end{aligned}$$

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

2.10.3 *Delay and echo*

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

2.10.4 *Crosstalk*

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

2.11 *Practical use of the model*

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

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

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

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

The usefulness of these facilities is evident.

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

3.1 *Introduction; type of model*

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

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

3.2 *Application to digital transmission*

3.2.1 *Definitions*

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

TABLE 3-1

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

Note 1 – Over the transmitted band.

Note 2 – At frequency f_j.

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

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

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

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

where M is the number of 16 ms segments.

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

3.2.2 *Basic formulas*

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

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

with

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

B_j is the weight allocated to the band of rank j; $C_j = 10 \log_{10}(f_j/\Delta f_c)$, Δf_c being the critical bandwidth.

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

TABLE 3-2

Frequency weighting

j	Equal articulation bands Extreme frequencies (Hz)		$B_j \times 10^5$	C_j (dB)
1	200	330	5 457	4.1
2	330	430	4 733	5.6
3	430	560	6 682	6.4
4	560	700	7 497	6.9
5	700	840	6 546	7.4
6	840	1 000	6 622	7.8
7	1 000	1 150	5 585	8.0
8	1 150	1 310	5 400	8.0
9	1 310	1 480	5 273	8.2
10	1 480	1 660	5 117	8.2
11	1 660	1 830	4 517	8.2
12	1 830	2 020	4 706	8.2
13	2 020	2 240	5 073	8.2
14	2 240	2 500	5 561	8.2
15	2 500	2 820	6 310	8.2
16	2 820	3 200	6 886	8.1
	TOTAL		102 158	

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

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

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

3.2.4 *Program and example of application*

The program used is reproduced in Appendix I.

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

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

3.3 *Application to analogue transmission*

3.3.1 *General; use of the program*

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

The program is used in the following way:

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

b) if I_c [5)] (information index under conversation conditions) is required, press "C"; read IN = I_c;

c) press "T".

3.3.2 *Data*

Lines 170-370 of the program.

Lines 180-370 correspond to ⅓ octave spaced frequencies from 0.1 to 8 kHz.

3.3.2.1 *Basic data*

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

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

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

BJ = frequency weighting (see [31]);

CJ = correction term in formula 3.4 giving V_j.

3.3.2.2 *Electroacoustic characteristics*

These depend on the connection considered.

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

D1 = Loss of the line filter.

---

5) See Table 3-5.

TABLE 3-3

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

System		MOS (0-4)	Over total band		Q_s in band No.																I_l (dB)
			Q_P	Q_{seg}	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
MNRU	15 dB	1.19	15	14.9	20.1	27.9	15.3	11.5	8.1	5,5	3.3	1.7	2.2	1.9	0.9	0.5	0.3	0.6	0.4	−0.6	17.1
	20	1.84	20	19.8	25.1	23	20.3	16.6	13.2	10.6	8.3	6.8	7.2	6.9	5.9	5.4	5.3	5.6	5.4	4.4	22.5
	25	2.44	25	24.9	30.4	28.5	25.6	21.3	17.5	15	12.6	10.8	11.4	11.1	10.1	9.6	9.5	9.8	9.4	8.2	25.2
	30	2.81	30	29.9	35.6	33.5	30.4	26.2	22.2	19.6	17.3	15.5	16.1	15.9	14.9	14.4	14.2	14.6	14.2	13	26.7
MNRUS	15	2.22	15.1	15.4	13.6	13.8	13.1	10.8	8.8	7.7	6.6	6	7.7	8.2	7.9	8.2	8.7	9.7	10.1	9.7	22.8
	20	2.59	20.1	20.4	18.6	18.9	18.1	15.7	13.6	12.5	11.5	10.8	12.5	13.1	12.8	13	13.6	14.7	15	14.6	25.8
	25	3.25	25.1	25.4	23.9	24.2	23.2	20.8	18.6	17.3	16.4	15.7	17.3	17.9	17.7	17.9	18.4	19.4	19.8	19.1	27.0
	30	3.31	30.1	30.4	29.1	29.3	28.2	25.8	23.4	22.2	21.2	20.5	22.1	22.8	22.5	22.7	23.1	24.1	24.5	23.9	27.4
F 16 kbit/s		0.81	14.2	11.4	15.2	13.5	10.8	7.3	4.1	1.7	−0.9	−2.1	−1.5	−1.9	−2.8	−3.3	−3.2	−2.5	−2.5	−2.7	16.6
24		1.59	20	17.7	21.7	19.8	17.4	13.9	10.6	8.1	5.9	4.7	5	4.5	3.8	3.5	3.1	3.6	3.4	2.8	23.4
32		2.59	25.2	22.8	27.1	25.4	22.6	19.2	16	13.7	11.4	10	10.3	9.8	9	8.6	8.4	8.8	8.5	7.9	26.0
V 16 kbit/s		2.03	12.8	14.5	16.2	14.2	11.9	8.6	6.1	4.2	2.5	1.9	2.6	2.2	2.1	2	2.4	2.8	2.9	2.8	24.3
24		3.06	19.7	21.8	25.1	23.2	20.6	17	14.1	11.8	9.3	8.2	8.8	8.3	7.3	7.1	7	7.7	7.7	7.2	26.7
32		3.28	25.7	28.3	32.1	30.1	27.6	24	20.5	18.3	15.6	14.4	14.8	14.3	13.3	13.1	12.8	13.3	13.2	12.3	27.4
SB 24 kbit/s		2.84	17.1	15	16.7	14.7	13	16.7	15.3	13.4	12.9	14.8	14.9	12.4	11.7	11.3	11.3	11.6	7.4	6.3	26.2
32		3.22	20.4	19.4	22.1	20	18.4	21.9	20.3	18.5	17.1	19.6	20	20.2	19.9	17.3	16.7	16.4	12.7	11.4	27.2

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

TABLE 3-4

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

Objective measure (Note)	Systems									
	Group B PCM, ADM, ADPCM-F		Group A same as B + ATC APC – AB		Group F ADPCM-F, ADPCM-V		Group E same as F + sub-band coding		Group D same as E + MNRU shaped MNRU	
	R	S	R	S	R	S	R	S	R	S
Q_P (SNR)	0.798	0.578	0.803	0.559	0.687	0.680	0.590	0.711	0.650	
Q_{seg} (SNR seg)	0.950	0.301	0.894	0.430	0.906	0.396	0.725	0.606	0.720	
"Log likelihood ratio"	0.943	0.213	0.924	0.341						
Cepstrale distance	0.954	0.208	0.929	0.331						
SRNF									0.884	
Information index					0.994	0.101	0.993	0.102	0.976	0.175

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

SNRF = Q_s frequency weighted

R = Correlation coefficient

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

3.3.2.3 *Noise components*

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

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

BDCN = spectrum of circuit noise;

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

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

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

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

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

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

EL = BDEL computed for RN = 50 dBA.

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

3.3.3 *Computation of signal-to-noise ratios*

3.3.3.1 *Level of the signal*

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

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

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

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

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

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

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

3.3.5 *Conversation information index, I_c*

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

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

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

3.3.6 *Examples*

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

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

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

TABLE 3-5

Information index I for NTT 600-type telephone sets (7 dB line) with SRAEN filter, STMR = 7.1 dB and opinion scores from tests 2 and 6

RN (dBA)	CN (dBmp)	ICN_0 (dBmp)	OLR (dB)	Y_L	I_L (dB)	Y_C	I_C (dB)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
60	−62.1	−58.2	1.4	3.13	23.48	2.94	23.27
			11.4	2.5	22.75	2.34	22.50
			21.4	2.31	19.44	1.58	19.52
			31.4	0.65	12.28	0.2	15.13
60	−59.8	−55.9	1.4	3.1	23.52		
			11.4	2.91	22.73		
			21.4	1.75	19.35		
			31.4	0.8	12.02		
60	−55.8	−51.9	1.4	2.83	23.59	2.99	23.38
			11.4	2.75	22.65	2.39	22.39
			21.4	1.79	19.03	1.28	19.20
			31.4	0.5	11.24	0.43	14.71
60	−51.4	−47.5	1.4	3.06	23.66	3.08	23.43
			11.4	2.24	22.44	2.17	22.16
			21.4	1.05	18.27	1.29	18.64
			31.4	0.09	9.3	0.22	14.11
60	−45.6	−41.7	1.4	2.31	23.64	2.63	23.37
			11.4	1.4	21.67	1.73	21.39
			21.4	0.64	16.2	0.77	17.23
			31.4	0.05	5.57	0.13	12.08

Explanation of columns

(1) Room noise, dBA

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

(3) ICN_0 = CN + 3.9 dB

(4) OLR (Rec. P.79)

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

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

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

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

TABLE 3-6

Information index at listening for NTT 600-type telephone sets (7 dB line) with SRAEN filter, STMR = 7.1 dB and listening opinion score from test 4

RN (dBA)	CN (dBmp)	ICN_0 (dBmp)	OLR (dB)	Y_L	I_L (dB)
0	(see Note)	−100 (see Note)	−3.6	2.30	22.87
			1.4	2.83	23.40
			6.4	3.26	23.85
			11.4	2.92	23.55
			16.4	2.59	23.05
			21.4	2.12	22.45
			26.4	1.89	21.73
			31.4	1.23	20.91
60	−55.8	−51.9	−3.6	2.61	22.78
			1.4	2.94	23.59
			6.4	3.00	23.51
			11.4	2.38	22.65
			16.4	1.80	21.24
			21.4	1.41	19.03
			26.4	0.91	15.95
			31.4	0.44	11.44
60	−56.9	−53	1.4	3.20	23.57
			11.4	2.53	22.68
			21.4	1.24	19.15
			31.4	0.24	11.51
50	−55.8	−51.9	1.4	3.21	23.88
			11.4	2.64	23.24
			21.4	1.58	21.04
			31.4	0.35	15.94
45	−64.9	−61	1.4	3.23	23.77
			13.4	2.62	23.24

Note – In these cases, there was no circuit noise during the opinion tests but a noise corresponding to CN = −76.9 (ICN_0 = −73) is used in the OPINE model. An arbitrary low noise value may be used for the calculation of the information index.

TABLE 3-7

Correlation between MOS and the information index in the case of analogue transmission

Type of connection	Range of conditions (see Note)			Type of MOS	Correlation coefficient	Deviation in terms of Y	
	RN (dBA)	ICN_0 (dBmp)	OLR (dB)			Standard error	Extreme deviations
NTT 600-type telephone sets + SRAEN filter	0 to 60	$-\infty$ and -61 to -52	-3.6 to $+31.4$	Y_L	0.978	0.15	-0.34 + 0.31
	60	-58.2 to -41.7	$+1.4$ to $+31.4$	Y_C	0.977	0.16	-0.32 + 0.32
13-2P-27-type telephone sets + attenuator	50	$-\infty$ and -71 to -49	$+4$ to $+48$	Y_C	0.995	0.07	-0.17 + 0.16

Note – Notations are of Appendix II.

3.4 *Possible extensions*

3.4.1 *Frequency characteristics*

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

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

3.4.2 *Connection including digital processes*

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

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

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

4.1 *Introduction*

NTT has been studying an objective model for evaluating telephone transmission performance [39], [40], [41], [42]. This describes OPINE (Overall Performance Index model for Network Evaluation), focussing on practical use.

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

OPINE was first constructed for fundamental factors in 1983 [39]. The opinion test data used for coefficient training and verification largely depend on the results of the experiment conducted at NTT ECL, Musashino in 1975. Its main purpose was to study the opinion score as a speech quality measure and a basis of telephone transmission standard. [40] and [43] describe the raw data. The experiment was of large scale with various factors taken into account, using an NTT 600-type telephone set.

In 1985, opinion tests were conducted for quantizing distortion. A newer revision of the model that also dealt with quantizing distortion was formulated and verified [41].

Some further opinion tests for talker echo and sidetone were conducted in parallel [44], [45]. A study of the evaluation characteristics of talker echo and its interaction with loudness was undertaken later.

In 1986 revision 2.0 of OPINE was formulated [43] in which all the parameters were rewritten in terms of loudness rating (LR). This revision was improved and updated to 2.1. Improved points in revision 2.1 are these minor changes:

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

While the model configuration was studied, the psychological characteristics of opinion evaluation were also investigated [46], using transmission loss and circuit noise as variables. The main conclusions were:

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

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

Therefore, we specify two classes of opinion tests:

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

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

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

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

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

4.2 *Outline of the model*

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

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

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

To calculate the PI for each factor, physical factors are obtained for loudness, distortion, etc., and each PI is transformed by an appropriate function. These functions are determined heuristically and the necessary constants are estimated from subjective data. The degree to which each factor influences the evaluation is reflected by these constants. The conceptual block diagram of OPINE is shown in Figure 4-1. The model consists of four parts: 1) an overall electro-acoustic calculation, 2) hearing parameter derivation, 3) a performance index derivation and 4) an evaluation derivation. The numbers in the figure refer to the equation numbers listed in § 4.3.

4.3 *Configuration of OPINE*

All the symbols are classified into 5 types:

Type [A]: model parameters

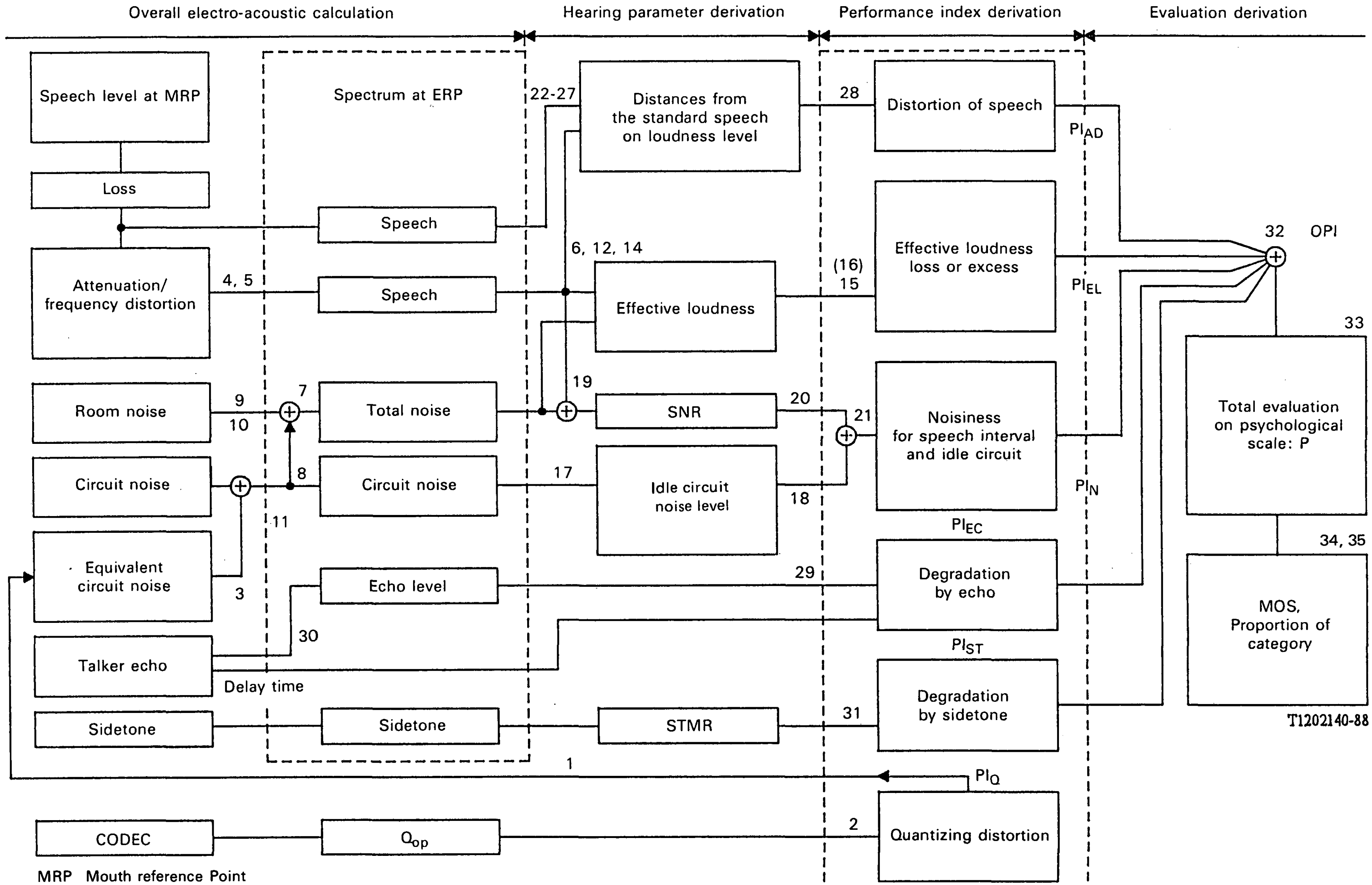
Type [A-1]: constants or coefficients adopted from standards

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

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

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

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



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

Note – The numbers in the Figure refer to equation numbers in § 4.3.

FIGURE 4-1

Block diagram of OPINE

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

4.3.1 *Overall electro-acoustic calculation*

4.3.1.1 *Opinion equivalent white noise level of quantizing distortion*

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

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

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

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

TABLE 4-1

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

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

where

(+) is the power summation operation

Type [B] symbols

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

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

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

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

Type [C] symbols

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

PI_Q is the PI for quantizing distortion.

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

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

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

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

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

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

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

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

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

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

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

Where:

(+) power summation operation

Type [A-1] Symbols

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

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

Type [A-2] Symbols

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

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

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

K_i is the critical bandwidth (dB)

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

Type [B] symbols

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

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

B_{RNi} is the room noise spectrum density (dB rel 20 µPa/Hz).
A-weighted evaluation of B_{RNi} becomes R_N (dBA)

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

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

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

Type [C] symbols

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

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

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

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

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

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

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

4.3.2 *Derivation of hearing parameters and performance index (PI)*

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

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

$$b_n = 44.38 \exp(-0.0869\ e_n) \tag{4-13}$$

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

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

where:

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

Type [A-1] symbols

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

Δf_i is the width of the ith frequency band (Hz)

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

M is the number of partitioned bands (= 19)

Type [A-3] symbols

λ_0 is the optimum loudness at ERP

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

Type [B] symbols

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

Type [C] symbols

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

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

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

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

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

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

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

where:

Type [A-3] symbol

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

Type [B] symbol

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

4.3.2.3 *PI_N (PI for noisiness)*

$$N_i' = \begin{cases} N'_{CQi} - N_{th}, & \text{when } N'_{CQi} \geqslant N_{th} \\ 0 \quad , & \text{when } N'_{CQi} < N_{th} \end{cases} \qquad (4\text{-}17)$$

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

where:

Type [A-1] symbol

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

Type [A-3] symbols

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

n is the exponent

Type [B] symbol

N'_{CQi} (see § 4.3.1.2)

Type [C] symbols

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

N_i' is the level above the noise threshold (dB).

$$SNR = 10 \log_{10} \left(\sum_{i=1}^{M} 10^{\frac{S_i}{10}} \Big/ \sum_{i=1}^{M} 10^{\frac{N_i}{10}} \right) \qquad (4\text{-}19)$$

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

$$PI_N = PI_{IN} + PI_{SN} \qquad (4\text{-}21)$$

Type [A-3] symbol

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

Type [B] symbols

S_i (see § 4.3.1.2)

N_i (see § 4.3.1.2)

Type [C] symbols

PI_{SN} is the PI for speech interval noisiness.

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

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

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

$$D_u = \sqrt{\frac{1}{M - M_s} \sum_{i=M_s+1}^{M} \Lambda_i^2} \tag{4-23}$$

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

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

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

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

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

where:

g_i is the conversion function from the speech power spectrum into a loudness level by equal-loudness curve (from [48])

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

Type [A-1] symbols

M is the number of partitioned bands (= 19)

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

Type [A-2] symbol

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

Type [A3] symbol

L_{th} is the loudness threshold (phon)

Λ_{th} is the threshold of Λi (phon)

Type [B] symbol

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

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

Type [C] symbols

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

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

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

D_u is the distance between Λ_l and Λ_d above 1 kHz

D_l is the distance between Λ_l and Λ_d below 1 kHz

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

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

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

$$E_0 \begin{cases} C_9 \log_{10} D + C_{10}, \text{ when } 0 < D < 60 \\ C_{11} \log_{10} D + C_{12}, \text{ when } D \geqslant 60 \end{cases} \tag{4-30}$$

where:

Type [B] symbols

E is the talker echo LR (dB)

D is the delay time of talker echo (msec)

Type [C] symbols

PI_{EC} is the performance index on talker echo

E_0 is the critical talker echo LR (dB).

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

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

where:

Type [A-3] symbol

St_0 is the critical STMR (dB)

Type [B] symbol

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

Type [C] symbol

PI_{ST} is the performance index on sidetone.

4.3.3 *Evaluation derivation* (see also Annex D)

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

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

where:

Type [A-3] symbol

P_0 is P with no degradation.

Type [C] symbols

OPI is the overall performance index

P is the mean overall evaluation on this psychological scale

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

or in practical form:

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

where:

Type [A-3] symbol

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

Type [C] symbols

MOS is the mean opinion score ranging from 0 to 4

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

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

Equations (4-34) and (4-35) are the adaptation of the model in [49].

4.4 *Symbol types and values*

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

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

TABLE 4-2

Input variables to the model

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

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

Note 2 St is calculated according to Recommendation P.79, § 8.

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

Note 4 R_N should be expanded B_{RNi}.

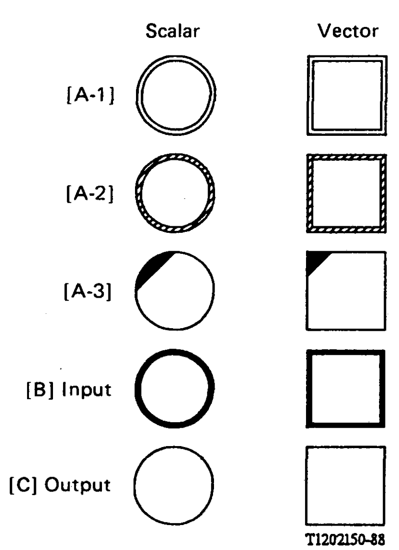


FIGURE 4-2

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

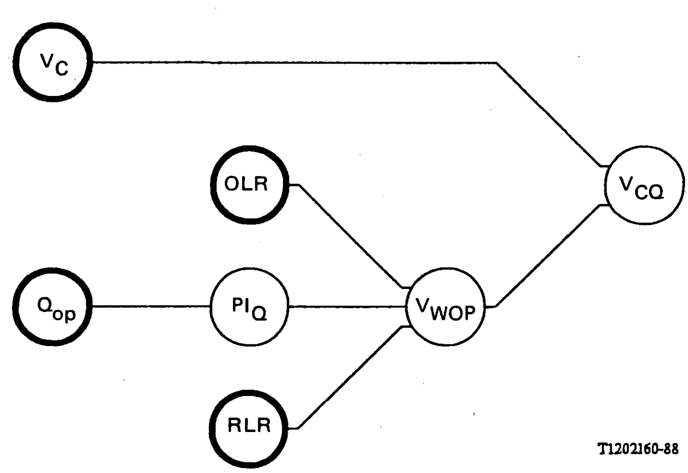


FIGURE 4-3

Block diagram of V_{CQ} calculation

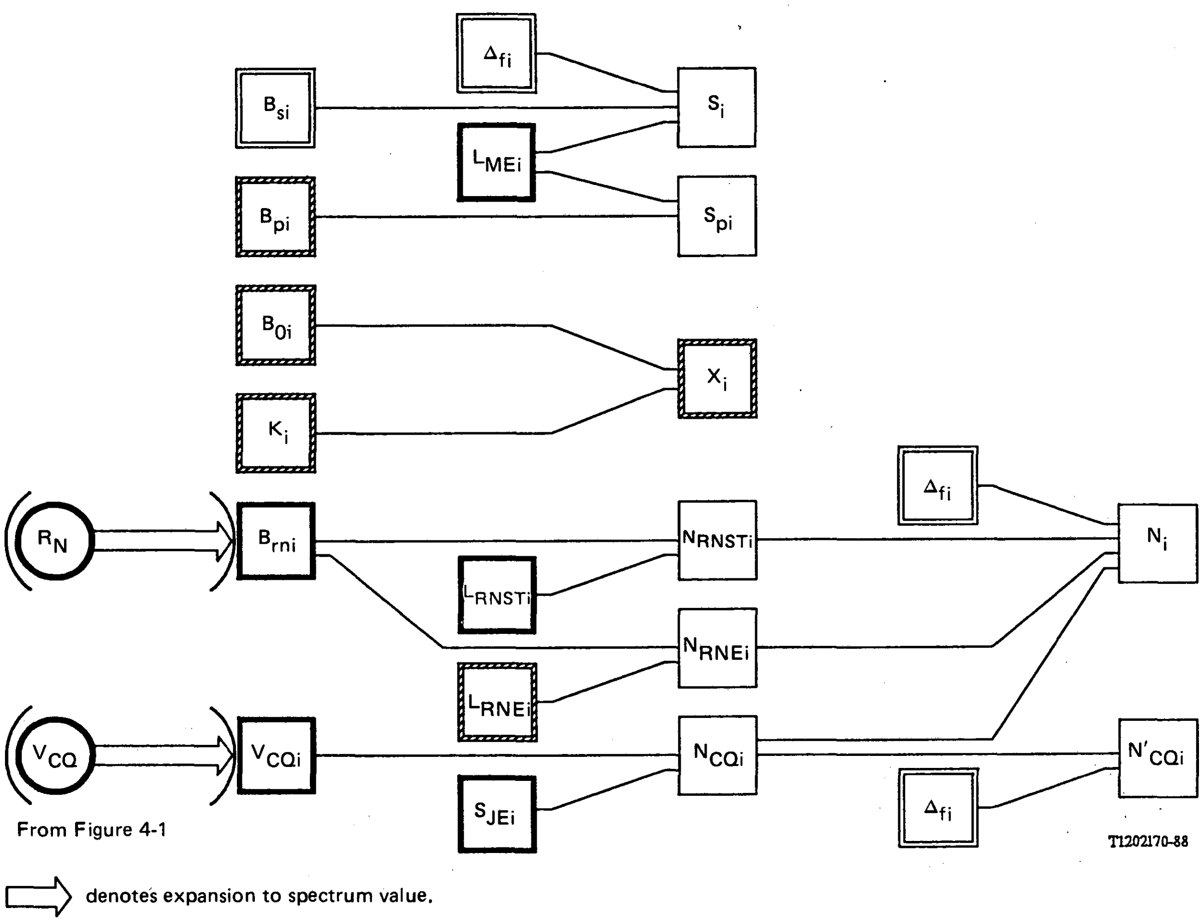


FIGURE 4-4

Block diagram of speech and noise spectrum calculation

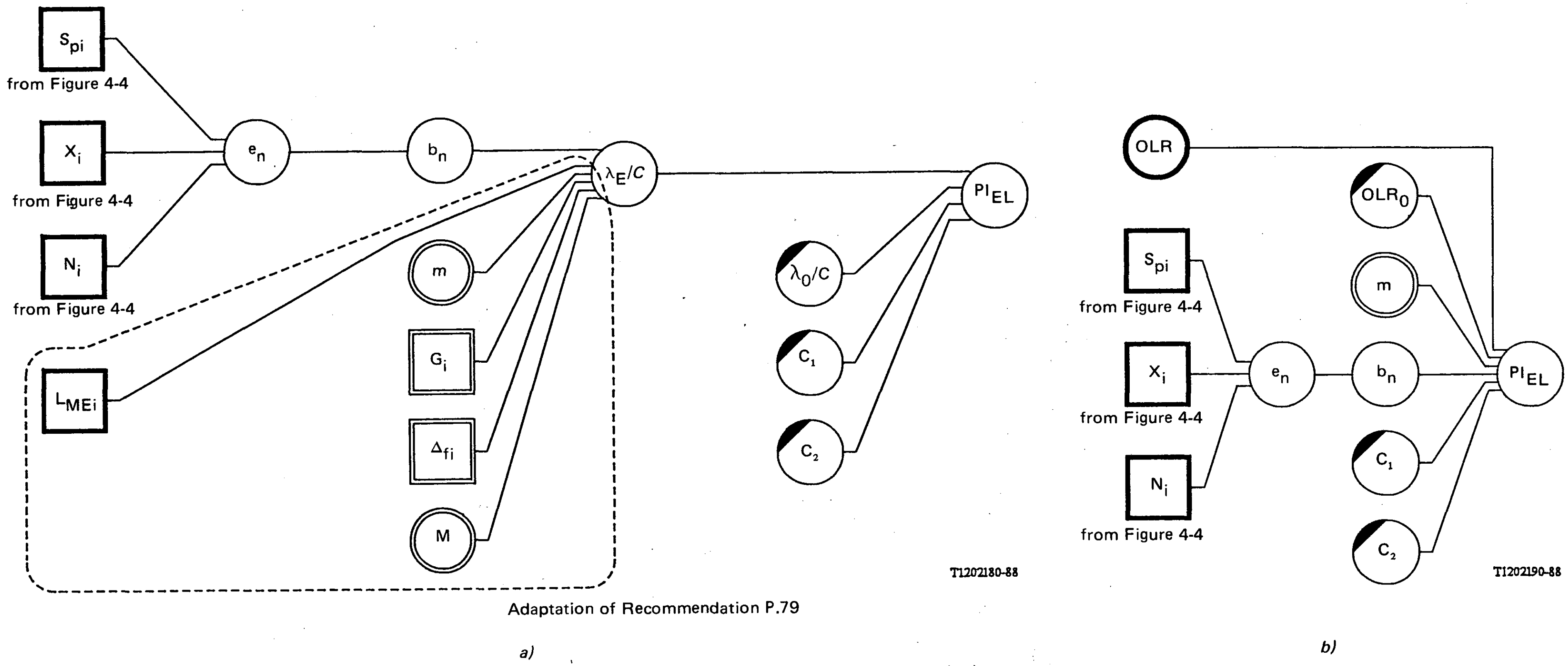


FIGURE 4-5

Block diagram of PI_{EL} calculation

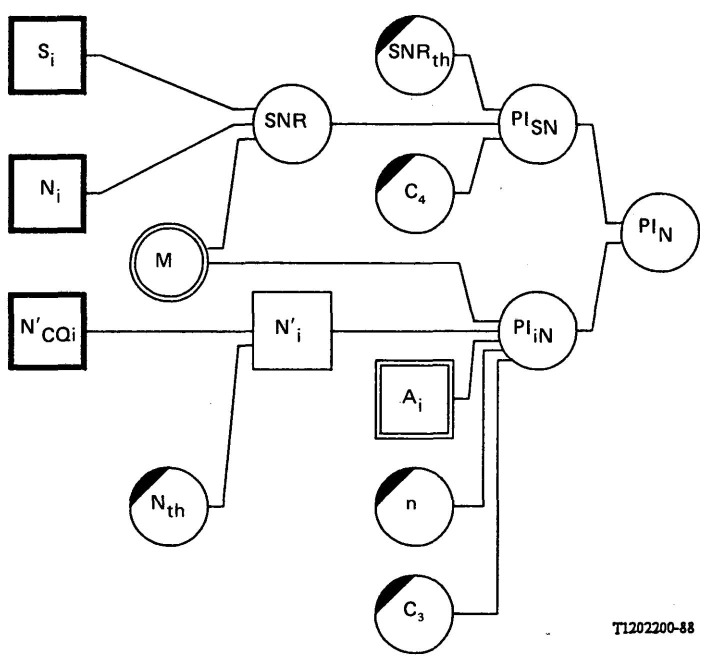


FIGURE 4-6

Block diagram of PI_N calculation

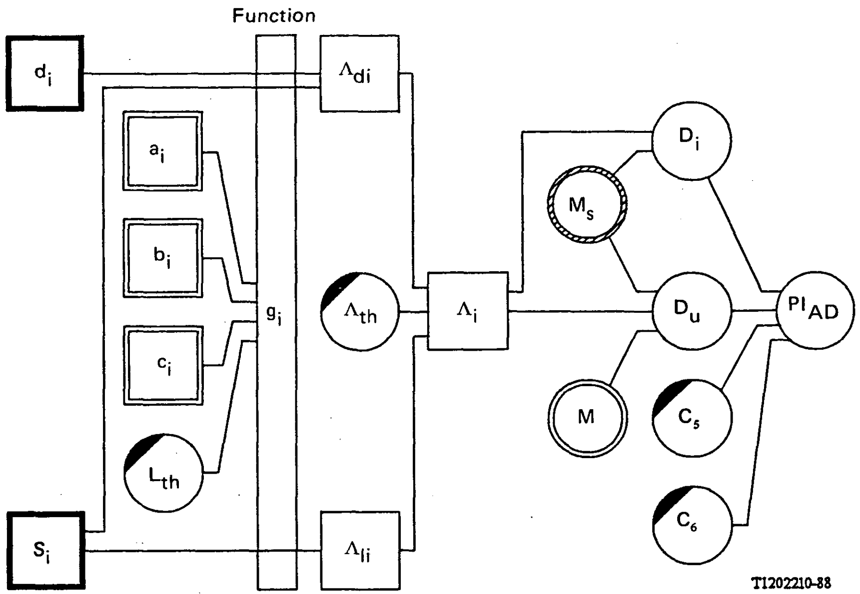


FIGURE 4-7

Block diagram of PI_{AD} calculation

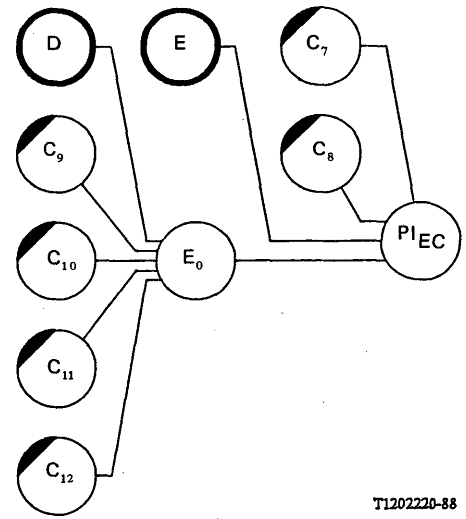


FIGURE 4-8

Block diagram of PI_{EC} calculation

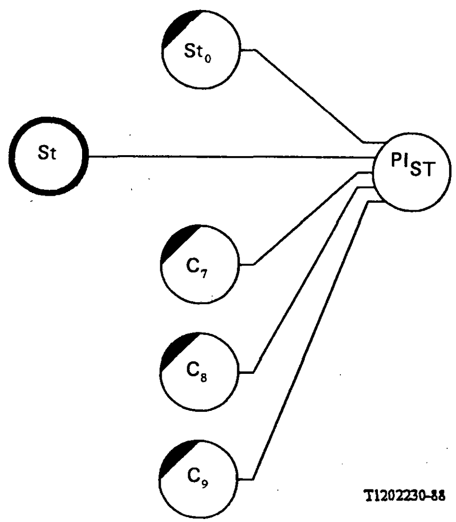


FIGURE 4-9

Block diagram of PI_{ST} calculation

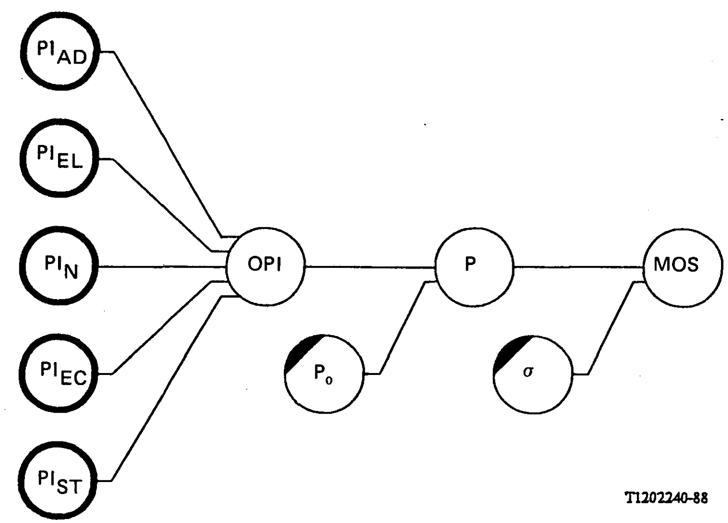


FIGURE 4-10

Block diagram of MOS calculation

TABLEAU 4-3

Values of a_i, b_i and c_i
(interpolated from [48])

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

TABLE 4-4

Model parameters

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

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

TABLE 4-5

Values of estimated constants and coefficients

No.	Related section	Output	Symbol	Value
1	4.3.2.1 4.3.2.2	PI_{EL}	C_1 C_2 λ_0/c OLR_0	0.0475 0.010 0.780 5.34
2	4.3.2.3	PI_{IN}	N_{th} n C_3	33.0 0.50 0.012
3	4.3.2.3	PI_{SN}	SNR_{th} C_4	7.5 −0.005
4	4.3.2.4	PI_{AD}	L_{th} C_5 C_6 Λ_{th}	57.5 0.043 0.043 15.0
5	4.3.2.5	PI_{EC}	C_7 C_8 C_9 C_{10} C_{11} C_{12}	13.69 0.01 26.4 2.65 14.00 24.6
6	4.3.2.6	PI_{ST}	C_{13} ST_0	0.00856 9.000
7	4.3.3	MOS	P_0 σ	3.558 0.730

ANNEX A

(to Supplement No. 3 – ref. to § 1.1)

Opinion ratings of transmission impairments

A.1 *Introduction*

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

A.2 *Overall loudness rating and circuit noise*

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

A.3 *Quantization noise from PCM processes*

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

A.4 *Bandwidth*

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

A.5 *Listener echo*

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

A.6 *Talker echo*

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

A.7 *Sidetone*

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

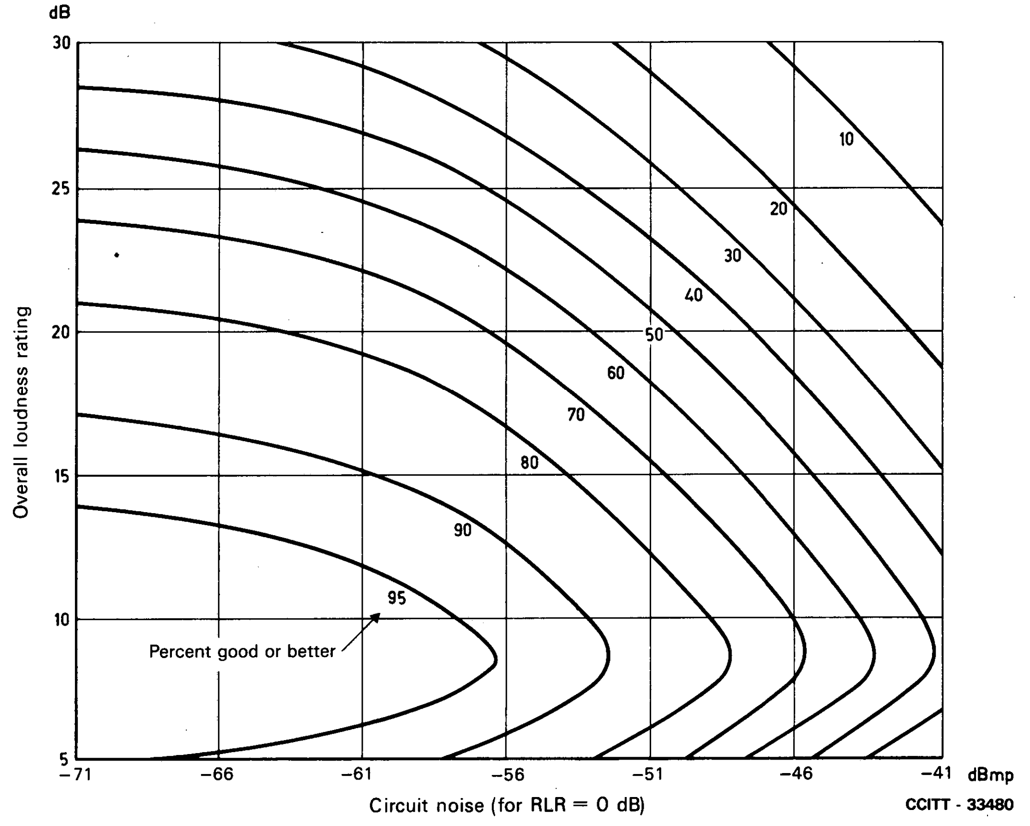


FIGURE A-1

Opinion rating for OLR and circuit noise

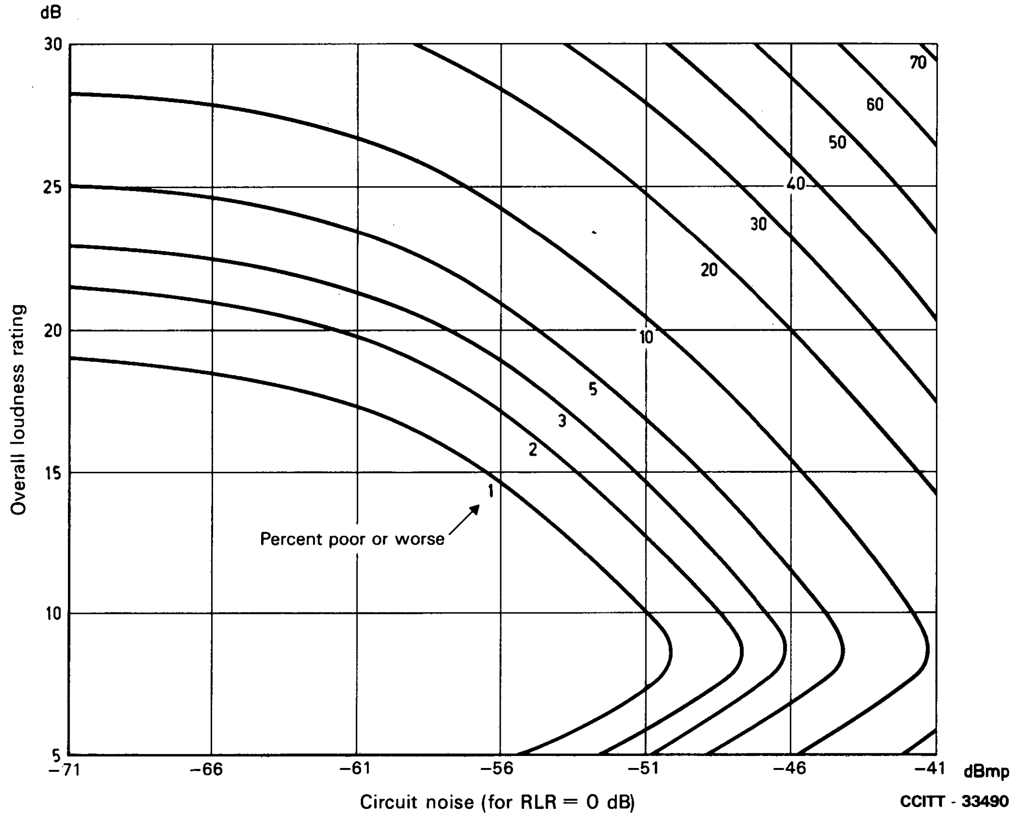


FIGURE A-2

Opinion rating for OLR and circuit noise

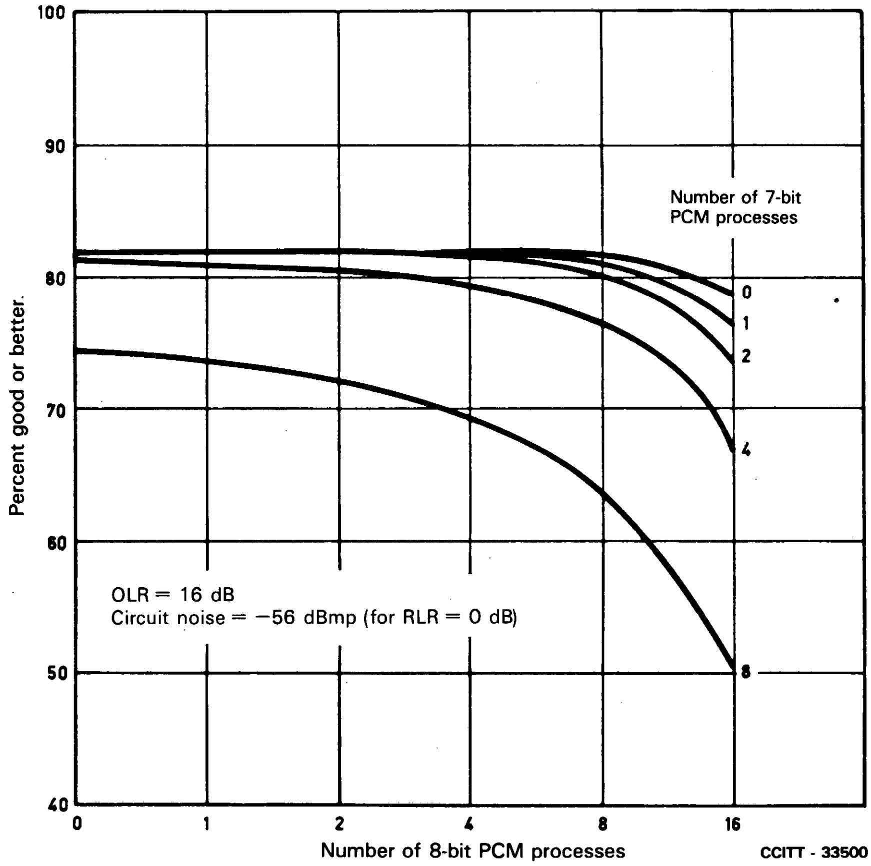


FIGURE A-3

Opinion rating for tandem PCM processes

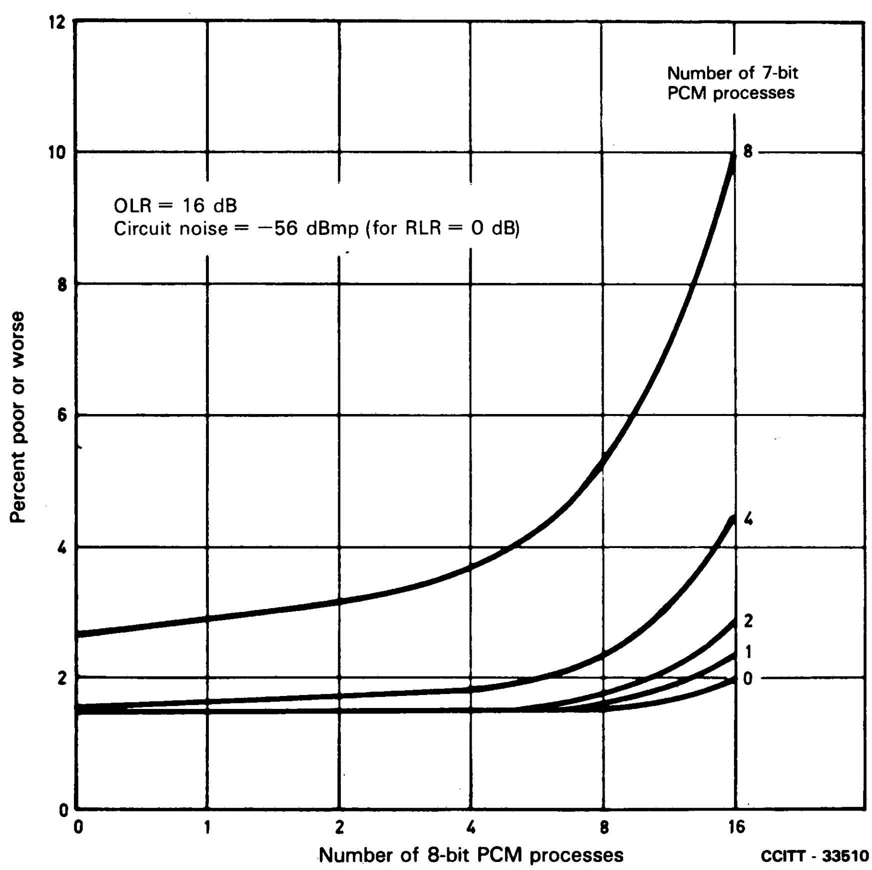


FIGURE A-4

Opinion rating for tandem PCM processes

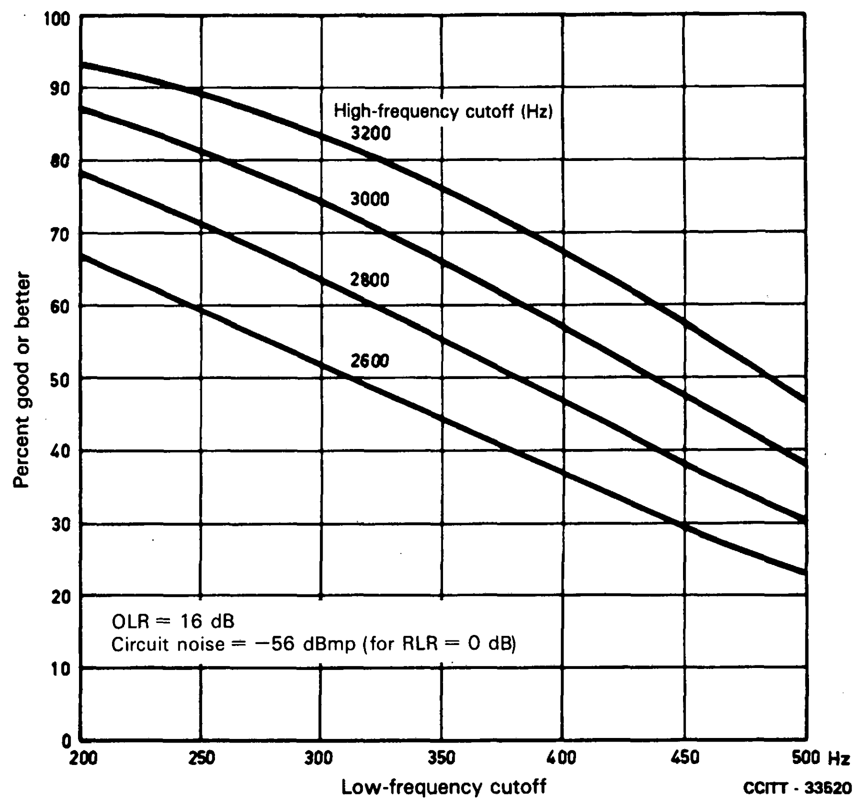


FIGURE A-5

Opinion rating for bandwidth

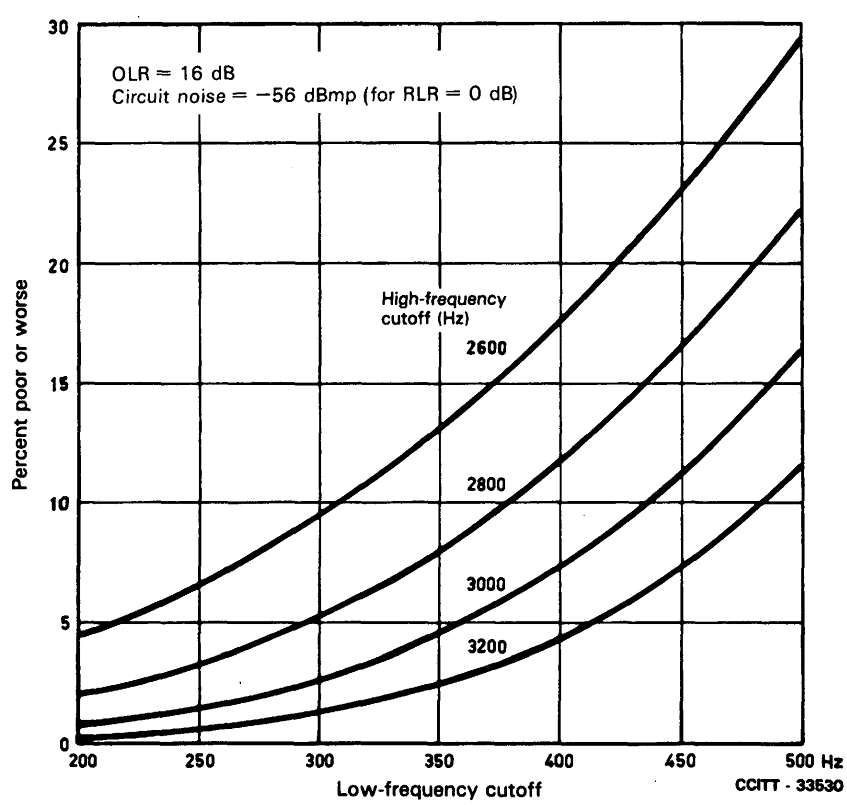


FIGURE A-6

Opinion rating for bandwidth

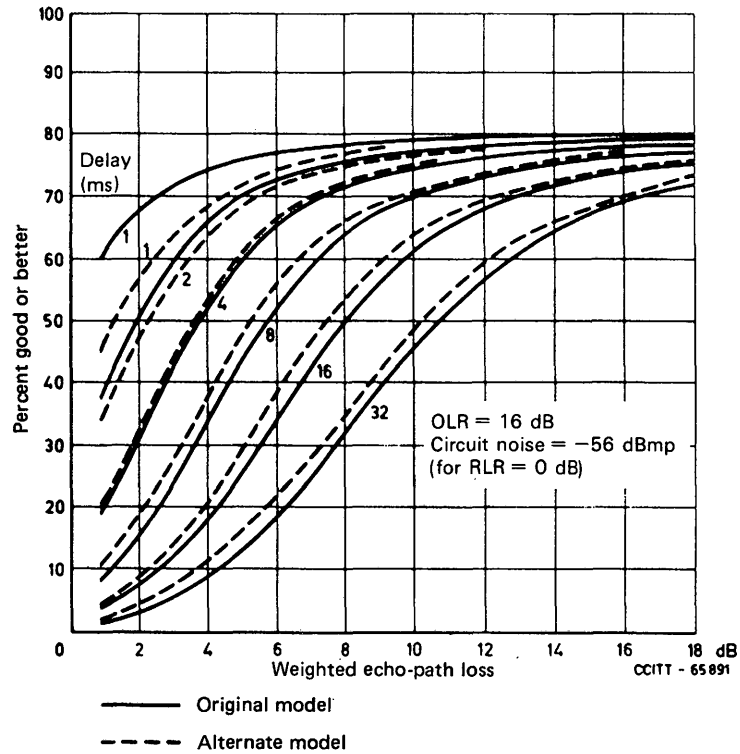


FIGURE A-7

Opinion rating for OLR, circuit noise and listener echo

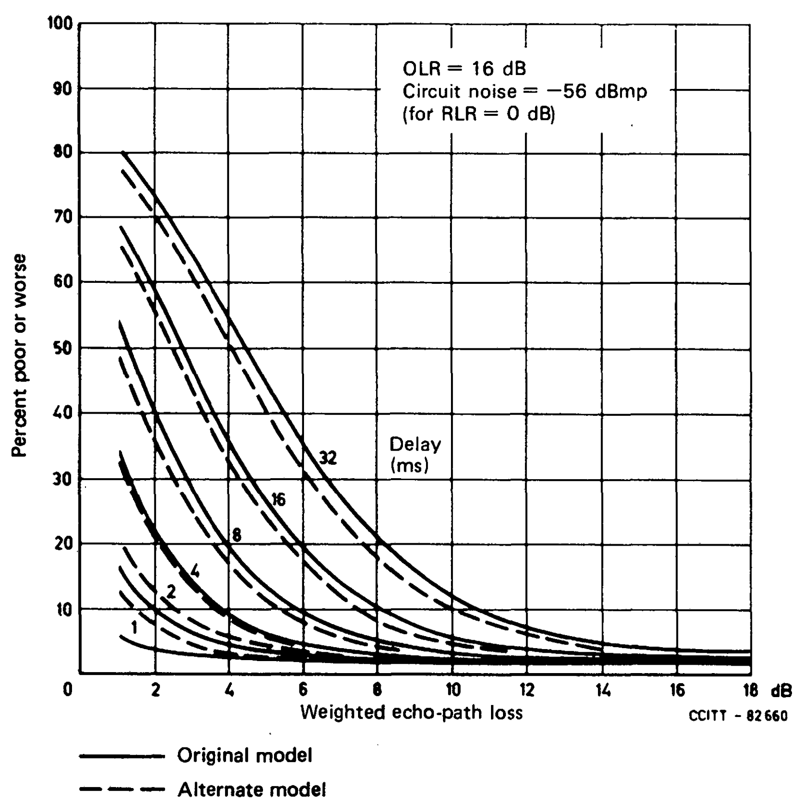


FIGURE A-8

Opinion rating for OLR, circuit noise and listener echo

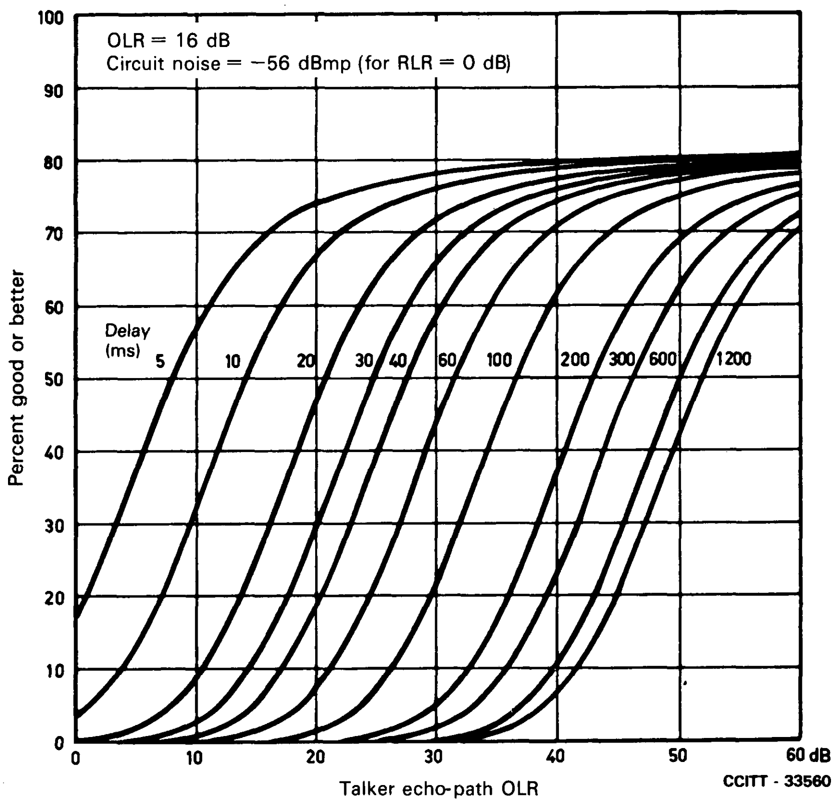


FIGURE A-9

Opinion rating for OLR, circuit noise and talker echo

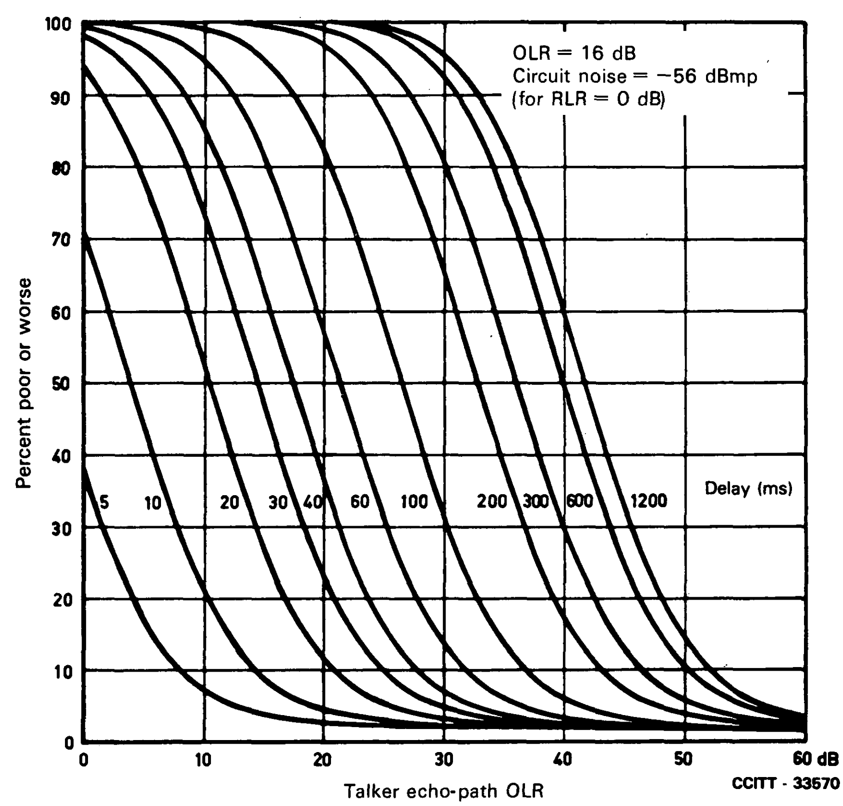


FIGURE A-10

Opinion rating for OLR, circuit noise and talker echo

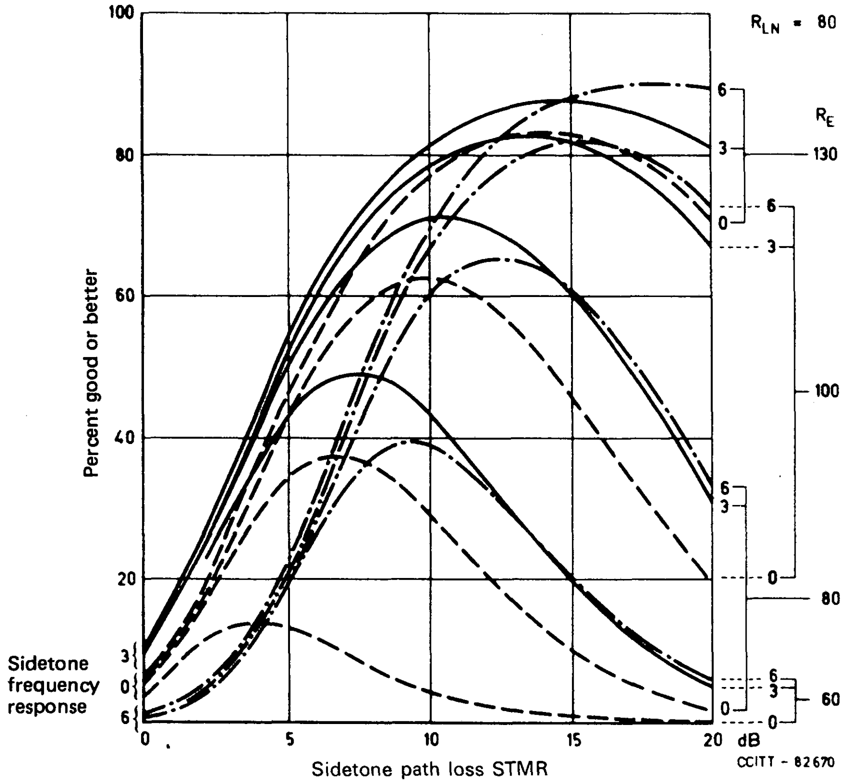


FIGURE A-11

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

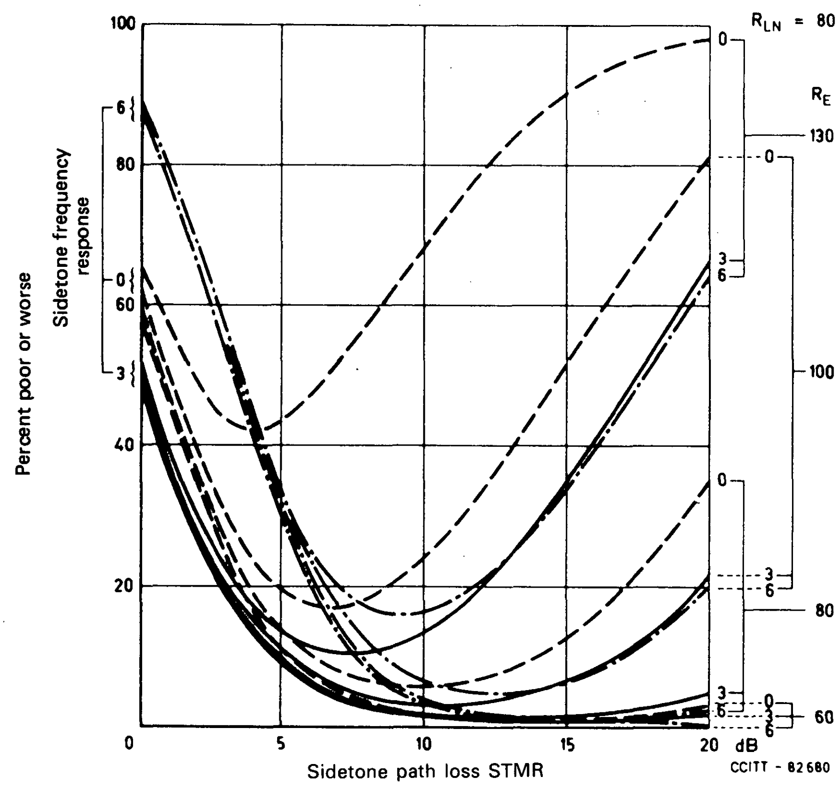


FIGURE A-12

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

ANNEX B

(to Supplement No. 3 – ref. to § 2.9)

Calculated transmission performance of telephone networks

B.1 *Introduction*

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

The loudness ratings quoted are calculated according to Recommendation P.79, using the frequency bands from 200 Hz to 4 kHz. The opinion scores, Y_{LE} and Y_C, are on a scale of 0 to 4, representing the listening effort and conversation opinion scales (see Supplement No. 2). The values of line current shown with the results are determined by the program which decides from the characteristics of the local telephone system which of a number of standard line currents is appropriate, and hence which values of the telephone instrument characteristics should be used. The program also gives speech levels for controlled talking conditions (V_L) and under conversational conditions (V_C). These and the loudness ratings are referred to the interfaces (NI and FI) shown in the figures below.

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

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

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

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

These two HRCs are both symmetrical and comprise BT limiting local lines of 1000Ω/10 dB, 4.5 dB local junctions and two 4-wire junctions each with 3.5 dB loss, which are the limits set by the BT transmission plan (given in [29]).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

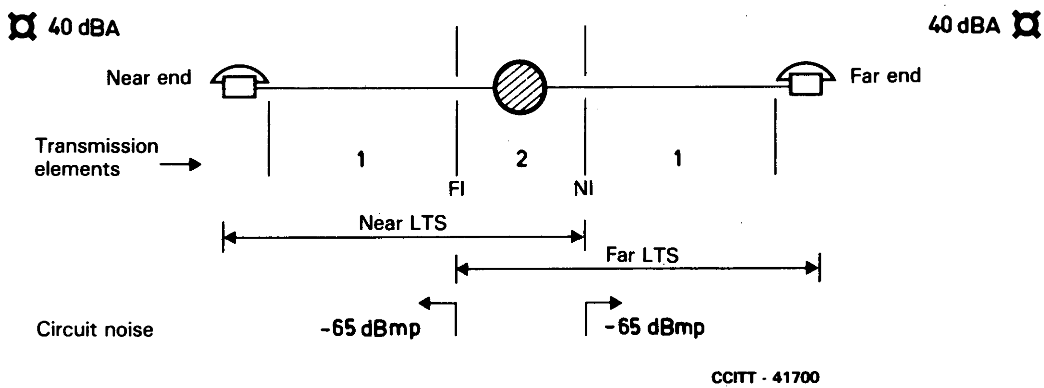
As a design tool, the program could be used further to verify the improvement in performance, to check the effects of tolerances and to consider possible improvements to these values.

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

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

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

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



Transmission elements

Telephone instruments are BT Type No. 706

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

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

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

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

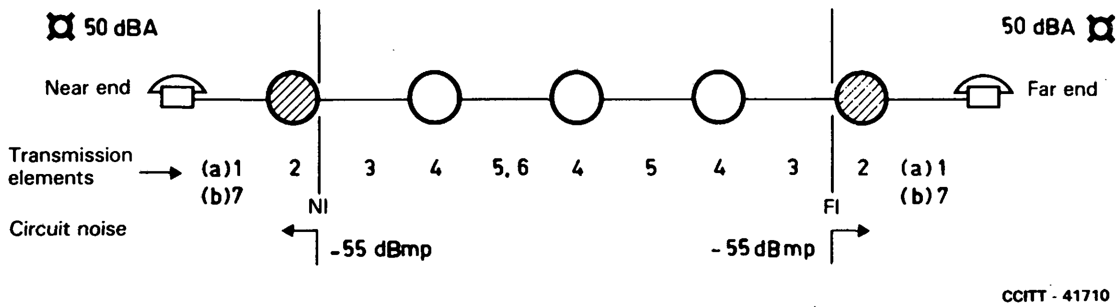
IL	Line current (mA)
SLR	Sending loudness rating (dB)
RLR	Receiving loudness rating (dB)
OLR	Overall loudness rating (dB)
STMR	Masked sidetone loudness rating (dB)
Y_{LE}	Listening effort score
Y_C	Conversation opinion score
V_L	Speech voltage at interface (dBV) under controlled talking conditions
V_C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

Note 1 – Room noise has Hoth spectrum.

Note 2 – The OLR printed in the left column is for near to far and the OLR in the right column is for far to near.

FIGURE B-1

HRC 1 Own exchange call



Transmission elements

Telephone instruments are BT Type No. 706

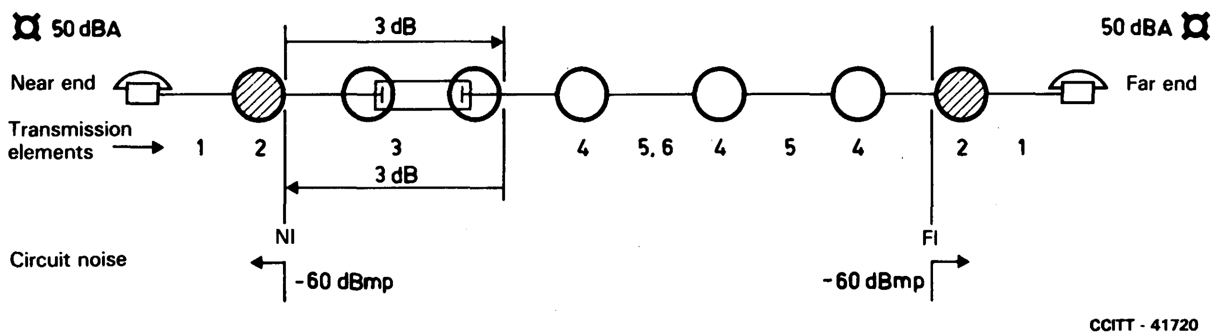
1 Unloaded cable 6 km of 0.5 mm (168 ohms/km, 50 nF/km)
2 Stone feed bridge (2 × 200 Ω, 2 + 2 μF, 50 V)
3 Loaded junction 19.6 km of 0.9 mm, 88 mH @1.83 km
4 Transformer feed bridge (50 V)
5 Attenuation 3.5 dB, frequency independent, 600 Ω
6 Channel filtering 300 Hz-3.4 kHz, 600 Ω
7 Unloaded cable 10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

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

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

FIGURE B-2

HRC 2 – Limiting national call



Transmission elements

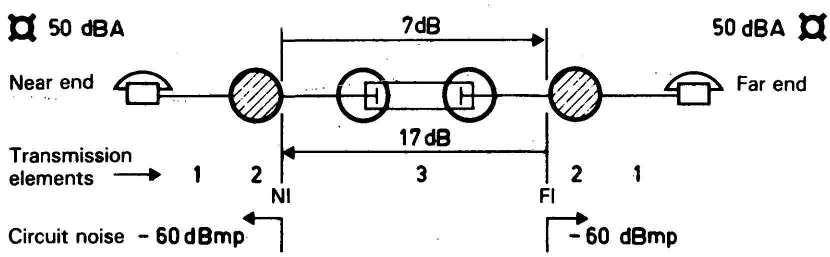
Telephone instruments are BT Type No. 706

1 Unloaded cable 1.6 km of 0.5 mm (168 ohms/km, 50 nF/km)
2 Stone feed bridge (2 × 200 Ω, 2 + 2 μF, 50 V)
3 PCM system 3 dB up to 3.4 kHz, 600 Ω
4 Transformer feed bridge (50 V)
5 Attenuation 3.5 dB, frequency independent, 600 Ω
6 Channel filtering 300 Hz-3.4 kHz, 600 Ω

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

FIGURE B-3

HRC 3 – Long distance call with a PCM junction



CCITT - 41730

Transmission elements

Telephone instruments are BT Type No. 706

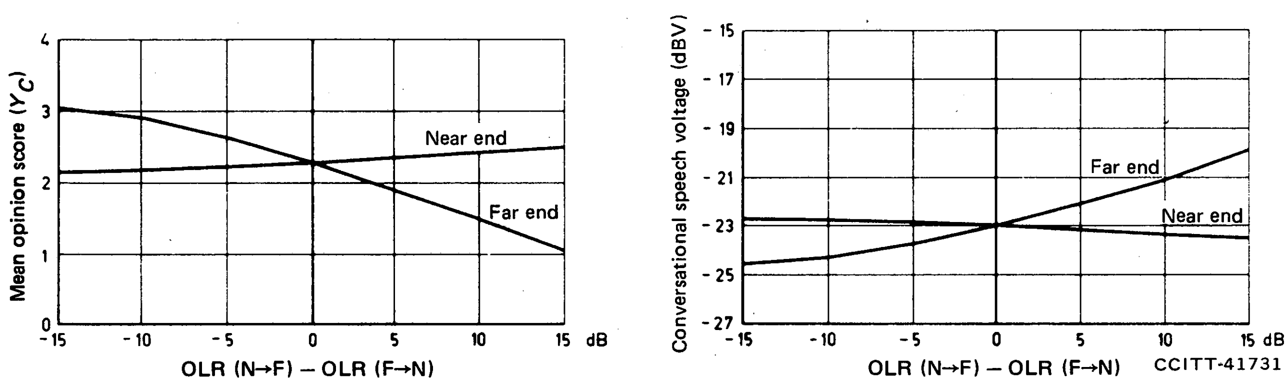
1 Unloaded cable 6 km of 0.5 mm (168 ohms/km, 50 nF/km)

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

3 FDM system loss as shown up to 3.4 kHz, 600 Ω

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

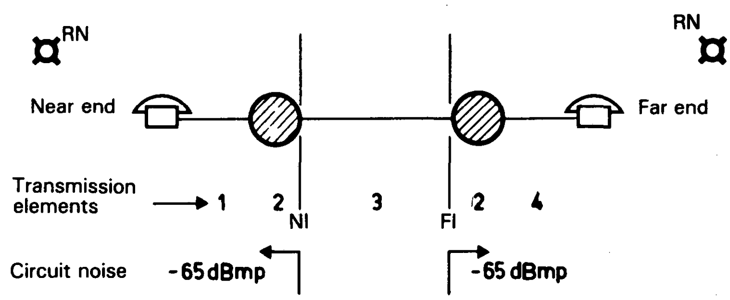
The results shown in the curves below are for the same connection with the loss from near to far in the FDM system varied from 2 dB to 32 dB. The loss from far to near was kept at 17 dB.



CCITT-41731

FIGURE B-4

HRC 4 – Effect of asymmetry of transmission loss



Transmission elements

Telephone instruments are BT Type No. 706

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

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

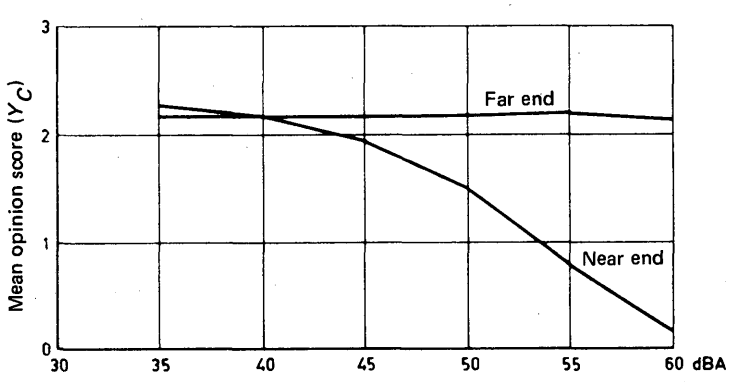
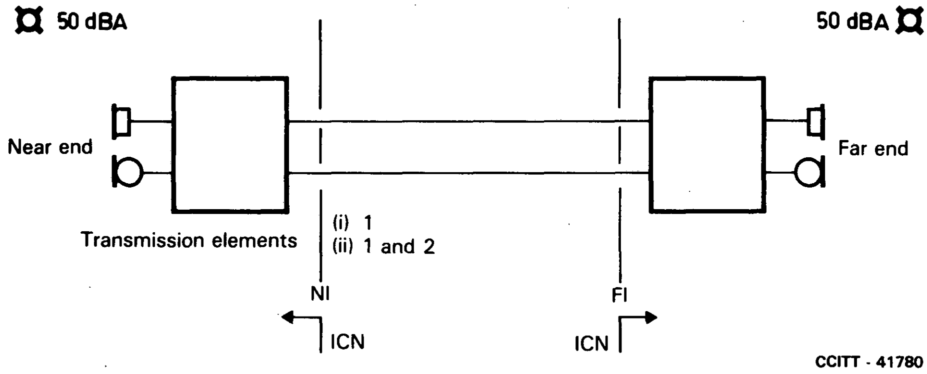


FIGURE B-5

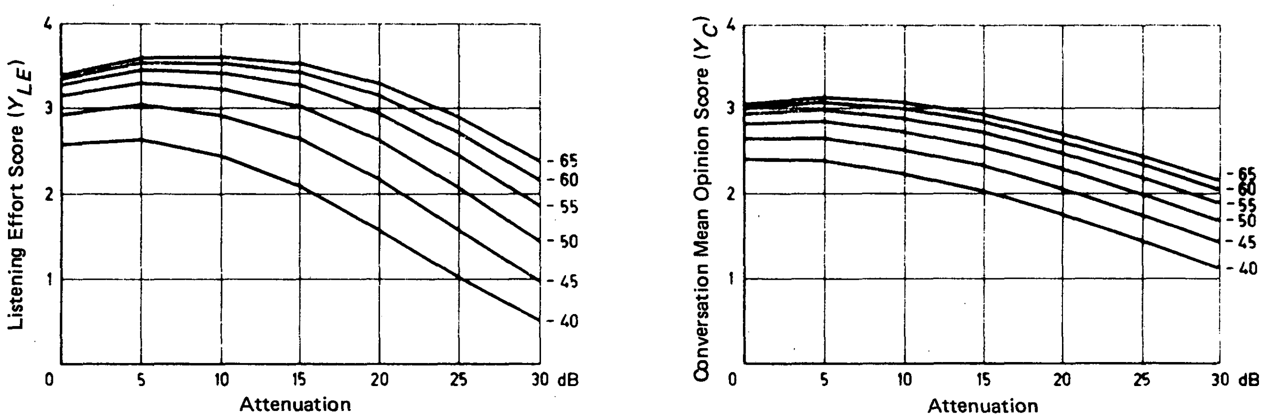
HRC 5 – Effect of room noise level



Transmission elements

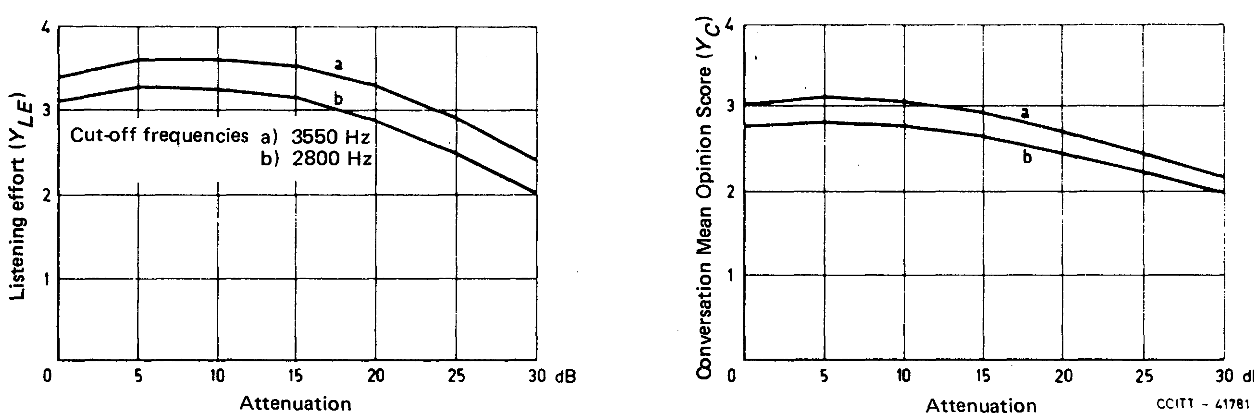
Telephone instruments are Intermediate Reference Systems (see Recommendation P.48) with a sidetone path of STMR = 20 dB.

1 Attenuation 0-30 dB, frequency independent, 600 Ω
2 Filtering 600 ohms, (a) 0-3.55 kHz
(b) 0-2.80 kHz



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

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

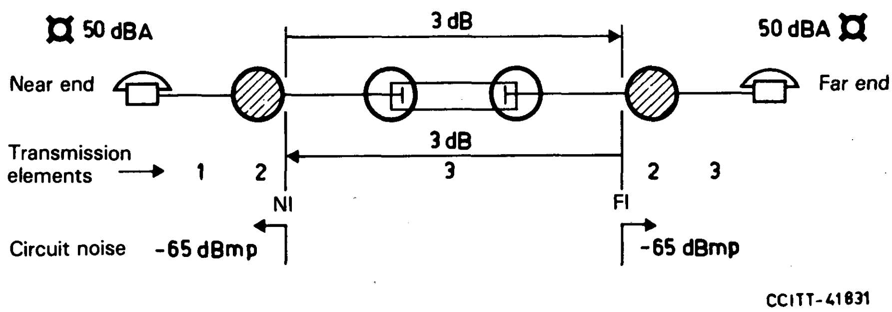


Note – These curves show the effect of bandlimiting with ideal lowpass filters.

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

FIGURE B-6

HRC 6 – Effect of injected circuit noise level and bandlimiting



Transmission elements

Telephone instruments are BT Type No. 706

1 Line: random selection from a sample of 1800 existing customers' lines.
2 Stone feed bridge (2 × 200 Ω, 2 + 2 μF, 50 V)
3 PCM system 600 Ω, 3 dB
4 Line: random selection of a line from the same sample of 1800 as in 1 above.

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

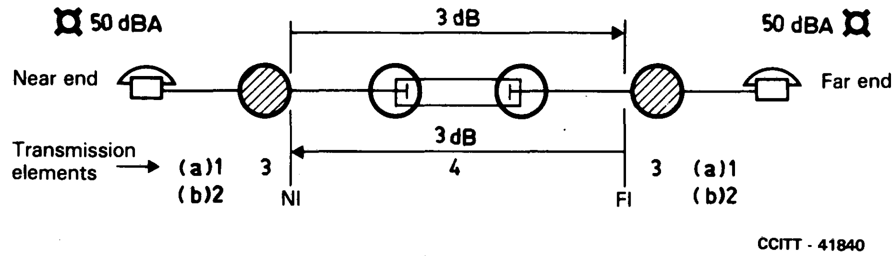
Line 731 (0.3 km)				Line 87 (0.5 km)			
Near end		*IL* =	75	*IL* =	64	*Far end*	
STMR =	7.19	SLR =	4.08	OLR =	2.42	RLR =	−5.41
RLR =	−5.74	OLR =	2.53	SLR =	4.50	STMR =	6.77
Y_{LE} =	3.45	V_L =	−17.46	V_L =	−18.16	Y_{LE} =	3.50
Y_C =	3.05	V_C =	−22.29	V_C =	−23.13	Y_C =	3.06
RN =	50.00	ICN =	−65.00	ICN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

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

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

FIGURE B-7

HRC 7 – Example with random selection of customers' lines



Transmission elements

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

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

Note – See also Tables B-1 and B-2.

FIGURE B-8

HRC 8 – Example of the use of CATNAP in design

TABLE B-1

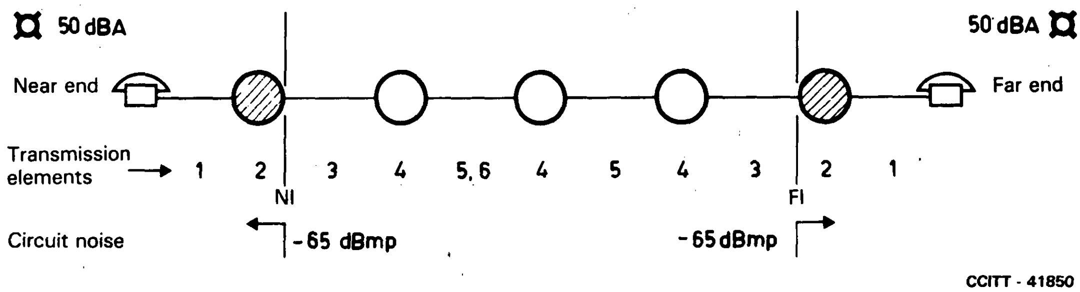
Values of STMR (dB) for specified lines (copper conudctors)

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

TABLE B-2

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

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



Transmission elements

Telephone instruments are BT Type No. 706

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

The results are shown in the curves.

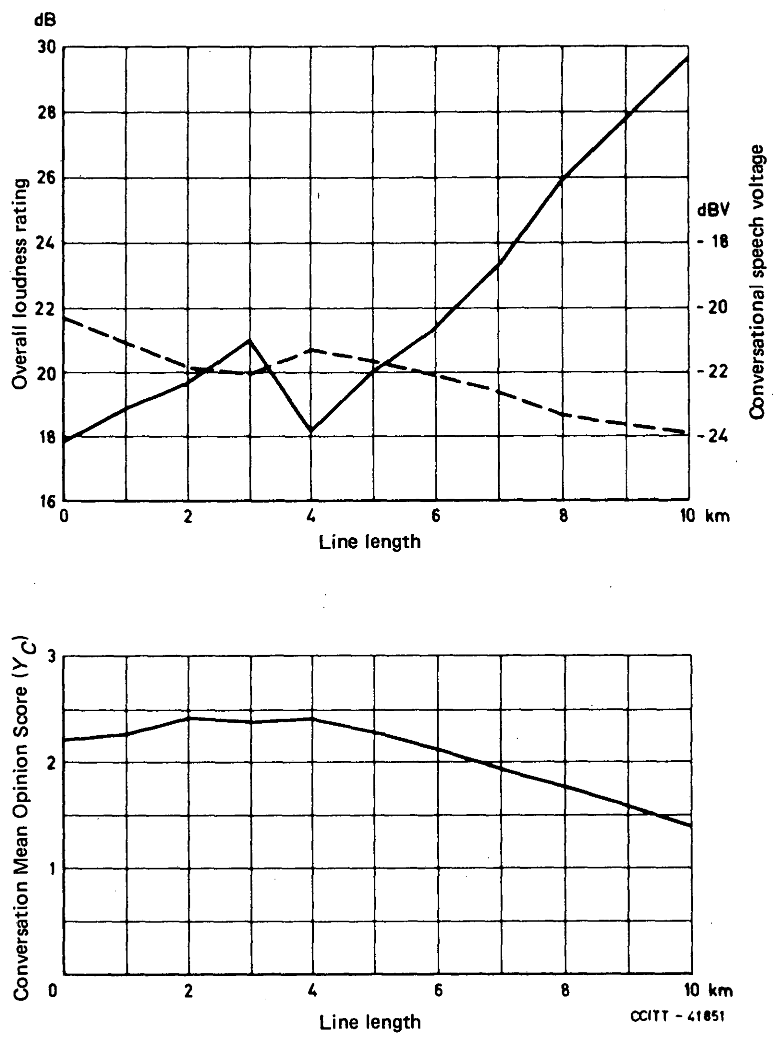


FIGURE B-9

HRC 9 – Effect of varying line length

ANNEX C

(to Supplement No. 3 – ref. to § 4.3.1.2)

Noise spectrum calculation

Expansion from the scalar value of noise to the spectrum values of both room noise and circuit noise is necessary (see Figure 4-4). The spectrum value database of R_N (60 dBA) and V_c (−56.0 dBmp) is shown in Table C-1. The value of room noise is taken from Figure 2/P.45 [50] and Figure 1 of Supplement No. 13. V_c is a mixture of circuit noise and switching office noise. They are expressed by flat noise and −8 dB/octave noise, respectively. If only a scalar noise level is known as a test condition, and its spectrum value is not known, then a mixed noise spectrum is used in OPINE in which −8 dB octave noise is 10 dB lower than flat noise. Moreover, SRAEN characteristics are added to the flat noise characteristics.

TABLE C-1

Noise spectrum value used in OPINE

R_N = 60 dBA V_C = −56.0 dBmp

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

ANNEX D

(to Supplement No. 3 – ref. to § 4.3.3)

MDS calculation examples

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

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

TABLE D-1

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

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

TABLE D-2

Example of estimated results for fundamental factors by OPINE

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

TABLE D-3

Attenuation/frequency characteristics used in Table D-2

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

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

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

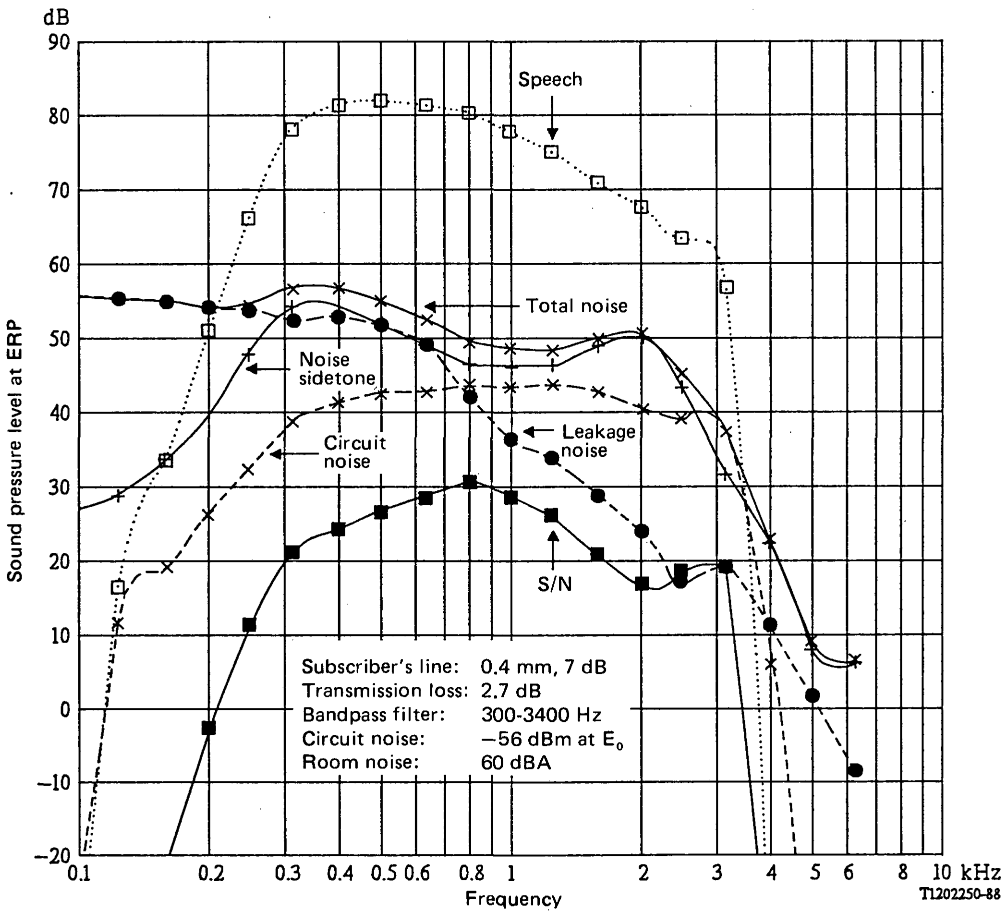


FIGURE D-1

Speech and noise level at ERP

ANNEX E

(to Supplement No. 3 – ref. to § 4.3.2.2)

Derivation of equation (4-16)

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

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

Taking the logarithm of equation (4-12),

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

Similarly,

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

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

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

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

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

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

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

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

ANNEX F

(to Supplement No. 3 – ref. to § 4.3.3)

Psychological evaluation model

This Annex gives a detailed derivation of equations (4-34) and (4-35). The model is a complete adaptation of [49].

F.1 *Psychological model for evaluation*

According to the model in reference [49], an evaluation value for a test condition on a psychological continuum is shown in Figure F-1. p_K is defined on page 10 of the reference, and is the probability of voting K as an opinion score for a test condition. The correspondences of opinion scores to ranges in the psychological continuum are:

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

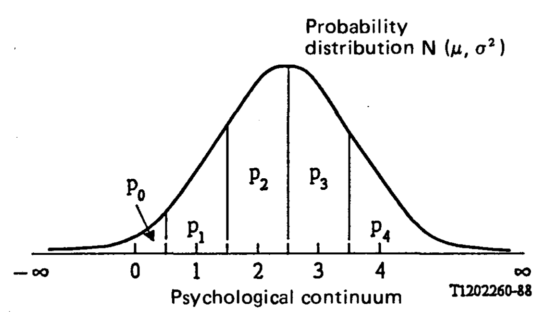


FIGURE F-1

Evaluation distribution for a test condition on a psychological continuum

These assumptions satisfy the following equation:

$$MOS = \sum_{k=0}^{4} k\, p_k \tag{F-1}$$

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

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

The cumulative probability of N (μ, σ^2) is expressed using a standard normal distribution function as follows:

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

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

$$MOS = \frac{1}{\sqrt{2\pi}} \left\{ 0 \times \int_{-\infty}^{(0.5-\mu)/\sigma} \exp(-t^2/2)\mathrm{d}t + 1 \times \int_{(0,5-\mu)/\sigma}^{(1.5-\mu)/\sigma} \exp(-t^2/2)\mathrm{d}t + 2 \times \int_{(1,5-\mu)/\sigma}^{(2.5-\mu)/\sigma} \exp(-t^2/2)\mathrm{d}t \right.$$
$$\left. + 3 \times \int_{(2.5-\mu)/\sigma}^{(3.5-\mu)/\sigma} \exp(-t^2/2)\mathrm{d}t + 4 \times \int_{(3.5-\mu)/\sigma}^{\infty} \exp(-t^2/2)\mathrm{d}t \right\} \tag{F-3}$$

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

$$MOS = \frac{1}{\sqrt{2\pi}} \left\{ \int_{(0.5-\mu)/\sigma}^{\infty} \exp(-t^2/2)\mathrm{d}t + \int_{(1.5-\mu)/\sigma}^{\infty} \exp(-t^2/2)\mathrm{d}t + \int_{(2.5-\mu)/\sigma}^{\infty} \exp(-t^2/2)\mathrm{d}t \right.$$
$$\left. + \int_{(3.5-\mu)/\sigma}^{\infty} \exp(-t^2/2)\mathrm{d}t \right\} \tag{F-4}$$

Since:

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

then:

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

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

APPENDIX I

(to Supplement No. 3 – reference to § 3.2.2)

```
10  PRINT "CALCULATION OF INFORMATION INDEX FOR CODECS AND MNRU"
20  REM New frequency weighting, Ti from BOSQUET, new equivalence with MNRU
30  REM PROGRAM ICQSKBE2, BAS, June 1987, written in MF BASIC
40  INPUT "SYSTEM"; S$
50  INPUT "MOS"; Y$
60  INPUT "K1(0 for MNRU, 5.2 in other cases)="; K1
70  DATA .05457, 4.1, .04733, 5.6, .06682, 6.4, .07497, 6.9, .06546, 7.4, .06622, 7.8, .05585, 8, .054, 8, .05273, 8.2,
    .05117, 8.2
80  DATA .04517, 8.2, .04706, 8.2, .05073, 8.2, .05561, 8.2, .0631, 8.2, 06886, 8.1
90  INPUT "QSEG over the band-QP=d(0 for MNRU and PCM)"; SM
100 REM calculation for codecs (for MNRU if K1=d=0)
110 FOR J=1 to 16
120 PRINT "Qseg over the band No"; J
130 INPUT QS
140 READ B, C
150 QC=QS+C
160 K2=1(1+EXP(-.159673*QC+.157246))
170 Q=K1+QC+K2*SM
180 V=3/(.1+10^(-Q/10))
190 I=B*V
200 II=II+I
210 NEXT J
220 REM Display of results
230 PRINT S$,"II="; II
240 LPRINT"        "; S$; TAB(20); K1; TAB(30); SM; TAB(40); II; TAB(50); Y$; TAB(60)
250 END
```

APPENDIX II

(to Supplement No. 3 – reference to §§ 3.2.2 and 3.3)

```
10   PRINT "Calculation of Information Index for NTT 600 sets (7 dB line)"
20   REM Program INT600E5 , written in MF Basic, September 1987
30   INPUT "Room noise,dBA="; RN
40   INPUT "STMR,dB="; STMR
50   INPUT "Circuit noise level (dBm, sign changed) at input to 0 dB
     RLR end"; I
60   ICNO=-I
70   INPUT "Listening (L) or conversation (C) or terminate (T)"; A$
80   IP A$="T" GOTO 640
90   IF A$="C" GOTO 560
100  INPUT "Overall loudness rating (P79),dB="; OLR
110  LPRINT "        OLR="; OLR
120  GOSUB 730
130  REM Correction for excessive loudness
140  IF OLR>OPT GOTO 380
150  x =2*OPT-OLR
160  GOTO 390
170  DIM FE (20), CN(20), ST(20), EL(20), BK(20), S(20), BJ(20), CJ(20), SRL(20), B1(20)
180  DATA -36.2,  -76.9,   -4.1,  32.4,  17.5, 56.0, .00804, 0   ,  62.3,  21.7
190  DATA -26.2,  -34.9,   -3 ,   31.2,  14.4, 61.1, .01042, 1.25,  27.8,  11.4
200  DATA -18.3,  -24 ,      .8,  29.5,  10 ,  62.5, .0138 , 2   ,  21.3,   6.9
210  DATA  -9.9,  -13.2,    5.6,  27.6,   5 ,  64.3, .01788, 2.6 ,  11.5,   5.6
220  DATA  -2.1,   -3.1,   12.7,  26.2,   2.5, 64 ,  .02392, 3.5 ,   3.1,   4
230  DATA   2.2,    6.9,   16.4,  22.3,   -.4, 60.7, .03246, 4.9 ,  -2.3,   2.1
240  DATA   3.9,   11.1,   16.6,  22.7,  -3 ,  59.8, .04471, 5.8 ,  -3.4,   1.2
250  DATA   3.2,   13 ,    13.5,  21.1,  -5 ,  59.4, .05981, 6.35,  -3.1,    .7
260  DATA    .8,   12.9,    8.9,  17.4,  -6.3, 56.3, .07789, 7.25,  -2.6,    .3
270  DATA    .3,   13.2,    6 ,    9.3,  -8 ,  52.4, .0839 , 7.35,  -2.5,   0
280  DATA   0 ,    12.4,    4.9,   2.7,  -9 ,  47.6, .0899 , 7.8 ,  -2.9,   0
290  DATA  -1.8,   11.5,    3.6,   -.9,  -8.5, 45.2, .09627, 8.05,  -2.8,   0
300  DATA   -.3,    9.7,    4.9,  -7.1,  -8 ,  44 ,  .10376, 8.25,  -2.1,   -.1
310  DATA   -.8,    6.4,    5.5, -12.4,  -9 ,  41.4, .11097, 8.3 ,  -4.6,   -.2
320  DATA  -9.9,    4.9,   -1.7, -20.4, -11.5, 38.8, .11859, 8.18,  -6.2,   -.2
330  DATA -17.1,   -2.4,  -15.4, -19.2, -13.8, 34.7, .12694, 7.95,   -.5,   0
340  DATA-111.4,  -54.7,  -24.8, -28.1, -13 ,  31 ,  .13607, 7.57,   7.9,  87.5
350  DATA-233.6, -144.8,  -40.4, -38.4, -12.5, 27.8, .14506, 7.25,  57.9, 200
360  DATA-237.2, -147.3,  -44.5, -51.3, -11.1, 26.1, .15487, 7.2 ,  62 ,  200
370  DATA-292.9, -199.8, -104.5, -66.6,  -9 ,  25.5, .16554, 6.8 ,  80 ,  200
380  X=OLR
390  DEF FNP (Y)=10 ^ (Y/10)
400  IN=0
```

```
410  FOR J=1 TO 20
420  READ FE, CN, ST, EL, BK, S, BJ, CJ, SRL, D1
425  REM Calculation and composition of signal to noise and equivalent ratio
430  PN=FNP(FE+RN-50-X+5)+FNP(CN+ICNO+60)+FNP(ST+RN-50-STMR+15)
     +FNP(EL+RN-50)
440  ZN=S+.4-SRL-D1-X+6.4-4.343*LOG(PN)
450  ZA=S+.8-SRL-D1-X-BK
460  IF ZA>0 THEN PE =(1+ZA/9.5)^2-1: GOTO 470
465  PE=1E-10
470  P=FNP (-ZN)+1/PE
480  Z=-4.343*LOG(P)
490  GOSUB 660
500  G=BJ*V
510  IN=IN+G
520  NEXT J
530  PRINT "IN="; IN
540  LPRINT "        RN(dBA)="; RN; "STMR(dB)="; STMR; "X(dB)="; X; "ICN0(dB)="; ICN0;
     "IN(dB)="; IN
550  GOTO 70
560  RESTORE
570  REM Speech power correction for sidetone and quality of conversation
580  IF STMR>13 THEN 590 ELSE 610
590  CS=0
600  GOTO 620
610  CS=.3*(STMR-13)
620  X=X-CS+.4085*IN-9.87
630  GOTO 390
640  END
650  REM Equivalence law and calculation of V
660  IF Z<1.74 THEN 670 ELSE 690
670  Q=Z+CJ
680  GOTO 700
690  Q=.494*Z+.88+CJ
700  V=3/(.1+10^(-Q/10))
710  RETURN
720  REM Determination of optimum OLR
730  RNS=RN-115+.006*(RN-30)^2-STMR-7.9
740  RNL=RN-121
750  PC=10^(ICN0/10)
760  PRL=10^(RNL/10)
770  PRS=10^(RNS/10)
780  N1=4.343*LOG(PC+PRL+PRS)
790  IF N1<-80 THEN OPT=7.2:RETURN
800  OPT=7.2-(N1+80)/8
810  RETURN
```

References

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

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

[3] CCITT – Question 4/XII, Contribution COM XII-No. R12, Study Period 1985-1988, Geneva, 1986.

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

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

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

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

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

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

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

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

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

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

[14] CCITT – Question 7/XII, Contribution COM XII-No. 1, Study Period 1985-1988, Geneva.

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

[16] CCITT – Question 7/XII, Annex 2, Contribution COM XII-No. 1, Study Period 1977-1980, Geneva, 1977.

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

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

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

[20] CCITT Draft Recommendation P.XXE, Question 15/XII, Annex 2 (II), Contribution COM XII-No.1, Study Period 1977-1980, Geneva, 1977.

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

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

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

[24] CCITT – Question 15/XII, Annex 1, § 3, Contribution COM XII-No. 1, Study Period 1981-1984, Geneva, 1981.

[25] CCITT – Question 19/X11, Contribution COM XII-No. 1, Study Period 1981-1984, Geneva, 1981.

[26] CCITT – Question 15/XII, Contribution COM XII-No. 1, Study Period 1977-1980, Geneva, 1977.

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

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

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

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

[31] LALOU (J.): a paper to be published in *Annales des Télécommunications.*

[32] CCITT – Question 7/XII, Contribution COM XII-No. 1, Study Period 1985-1988.

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

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

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

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

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

[38] CCITT-Contribution COM XII-R.17, Report of WP XII/3 meeting in Budapest, May 1987 (Reply to Question 4/XII).

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

[40] CCITT-Contribution COM XII-235, Calculation method of OPINE, (NTT), Geneva 1984.

[41] CCITT-Contribution COM XII-10, Objective evaluation model of telephone transmission performance for fundamental transmission factors and quantizing distortion, (NTT), Geneva 1985.

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

[43] NTT: OPINE (Rev. 2.0), private communication, Sept. 1986

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

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

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

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

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

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

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

CONSIDERATIONS RELATING TO TRANSMISSION CHARACTERISTICS FOR ANALOGUE HANDSET TELEPHONES

(Malaga-Torremolinos, 1984; amended Melbourne, 1988)

1 Introduction

This Supplement based on reference [9] summarizes available information on how some characteristics for handset telephones can be optimized.

It contains information about sending and receiving sensitivities, frequency responses, sidetone characteristics, influence of impedance and handset dimensions. It must be remembered that there are different ways to make an optimization. For instance the number of degrees of freedom are essential. As there are different opinions in different countries (for instance, the different assumptions made) the results of the optimization will be different. This Supplement touches some of these aspects.

2 Receiving frequency response

Most Administrations seem to prefer a fairly flat frequency response between 300 Hz and 3400 Hz. This probably derives from the early days of telephone networks, when it was determined that possible pre-emphasis at higher frequencies should be located at the sending end to obtain the best possible overall signal-to-noise performance. If we consider free-field, two-ear listening as a reference (face to face conversation) and assume a frequency-independent (flat) response, we should in principle simulate these conditions also at one-ear telephone listening.

Then, at the earphone listening, we should have a frequency response of the earphone as in Figure 1 to simulate the diffraction effect we have at free-field two-ear listening [1]. However, most Administrations seem to prefer a flat response and to put the corresponding correction at the sending end. It may also be easier to construct a receiver with high efficiency if the goal is a flat response. Reference [2] has suggested a response as in Figure 2 optimized for a mean local line. Where mains noise may cause problems, a response with greater loss at lower frequencies, e.g. at 200 Hz and lower frequencies, may be appropriate.

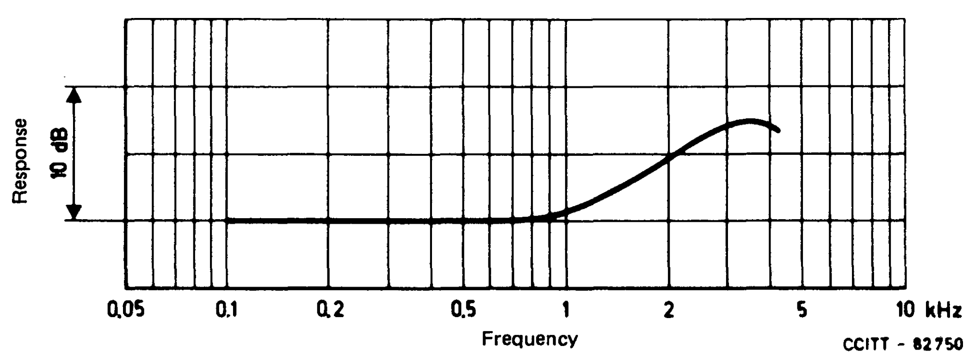


FIGURE 1

The diffraction effect around the head at 1 m distance in free field [1]

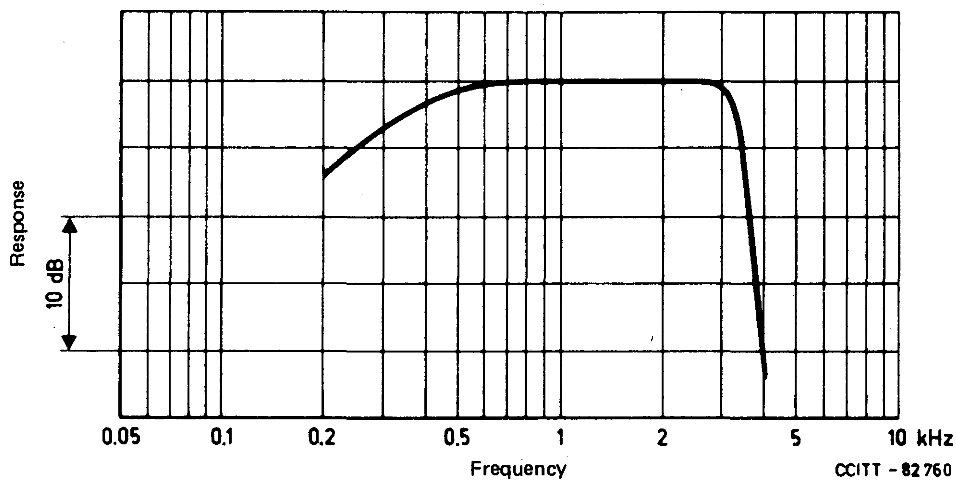


FIGURE 2

Receiving frequency response according to [2]

3 Receiving sensitivity

Receiving sensitivity today often is represented by values between an RLR of −4 dB and −12 dB respectively.

A further increase of the sensitivity by the use of amplifiers might technically be possible. However, the probability for the audibility of crosstalk will increase with increased sensitivity. Therefore, the information gathered in Recommendation P.16 must be considered and it is doubtful if it can be recommended to increase the sensitivity further beyond an RLR of −12 dB.

Increasing the receiving sensitivity also decreases the margins against the effects of speech-off noise on the connection, e.g. unwanted modulation products from PCM systems. The stability against singing will also be affected.

4 Sending frequency response

Having chosen the receiving response to be flat, the sending frequency response can be optimized to give the proper overall characteristic. Reference [3] suggests an optimization achieved by asking the listeners for the "preferred" response. The result is shown in Figure 3. Reference [4] suggests a 2 to 3 dB increase per octave with increasing frequency. This result was obtained in tests regarding "naturalness". Reference [2] suggests a steeper curve (Figure 4) as a result of an optimization where maximum loudness, minimum listening effort and lowest output level are combined. The degree of freedom used by [2] is of course less than in [3] and [4]. Here we may have a difference in opinion concerning which assumptions we must include in the optimization. If the signal-to-noise ratio is a problem, some decibels could be gained (without overloading) in the way shown by [2]. If there are no signal-to-noise ratio problems, an optimization for best naturalness as in [3] and [4] can be used. Thus, the result will depend on the assumptions.

Different opinions may also exist about the local cable length for which the frequency response should be optimized and if the high frequency loss at long lines should be compensated. Reference [2] suggests optimization of the mean local line which will be optimum to the highest number of subscribers (because of the statistical distribution of cable lengths).

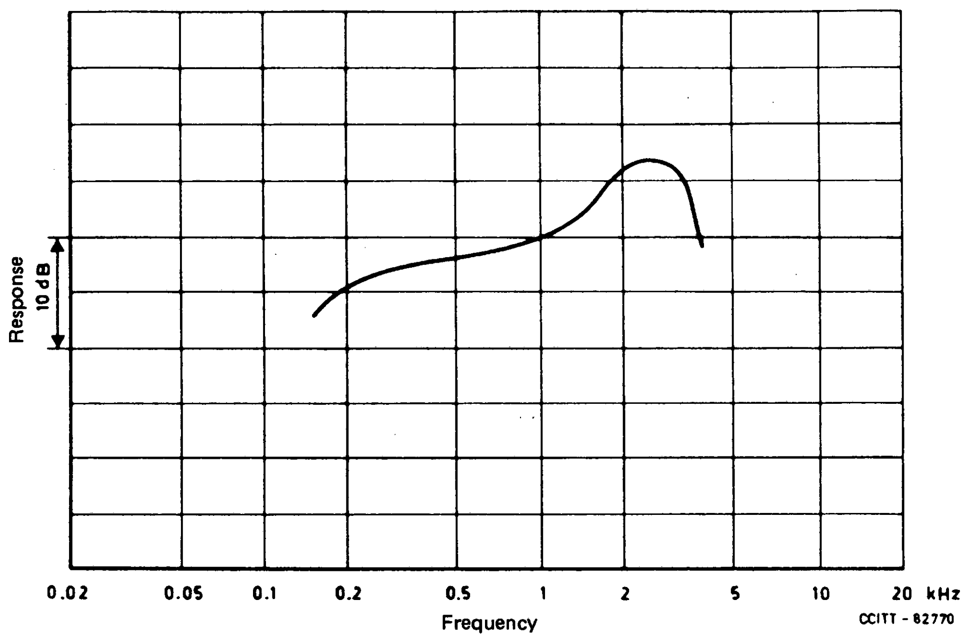


FIGURE 3

Sending frequency response according to [3]

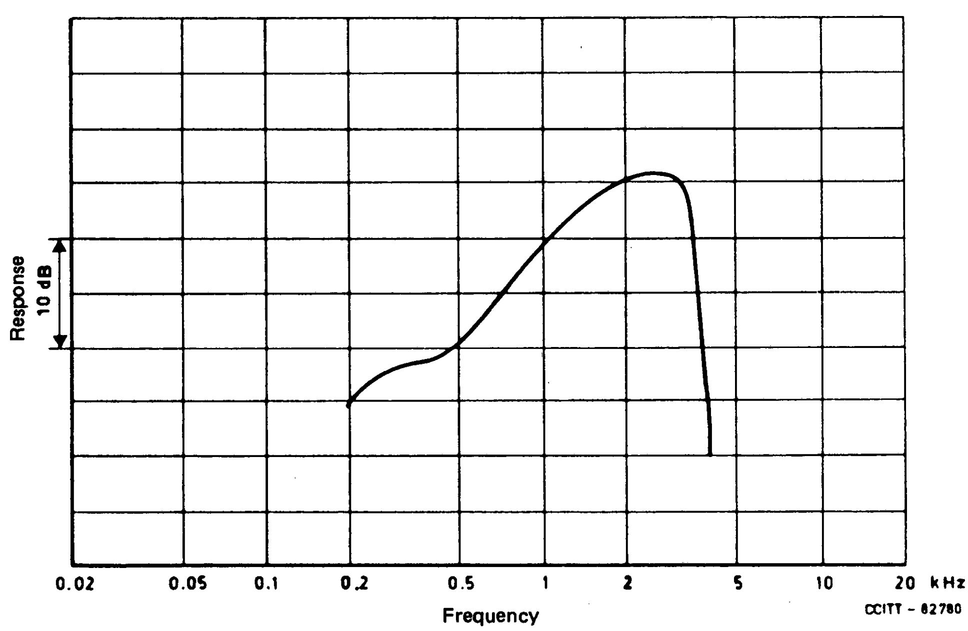


FIGURE 4

Sending frequency response according to [2]

The curves according to Figure 4 and [4] give with a flat receiving frequency response an overall characteristic close to what is obtained by the diffraction effect at free-field listening. However, this is probably not the whole explanation to the preferred curves. Even if the receiving responses were flat during sealed measuring conditions, hardly anyone keeps the earphone tight to the ear during conversation. Therefore, the actual responses during conversation probably give some additional low frequency cut-off that certainly has an influence on the results (see also reference [5]).

5 Sending sensitivity

When we want to choose the sending sensitivity we have one degree of freedom less that at the receiving end. We must consider both the probability of crosstalk and the probability of overloading other parts of the telephone system. Actual output levels from the telephone must be considered. As shown in [6] different output levels for the same SRE-value have been found in different countries. However, the different results show one important feature in common: output levels during normal conversation are generally lower than during reference equivalent measurements. Hopefully we will get better agreement on this point in the future if we use the measuring distance defined in Recommendation P.76, Annex A for loudness rating measurements.

6 Regulation

A possibility to increase the sending sensitivity on long lines exists if we use sending regulation dependent on line length. The probability for overloading and the probability for far end crosstalk will not increase if the mean power is kept to the same value as today. See also [2]. The probability of near end crosstalk in the local cable will of course increase and has to be considered.

If regulation is introduced both at sending and receiving, more subscribers may experience an overall loudness rating close to a preferred optimum, i.e. less calls will be rated poor and unsatisfactory. Another reason to introduce regulation is to obtain a better sidetone performance on short and long lines at the same time.

7 Impedance presented to the line

Some considerations concerning this topic are as follows:

- a conjugate match with the line maximizes the power transferred but creates sidetone problems on short lines and also stability/echo problems on long-distance calls;
- an image match to the line reduces the range of impedance presented to the exchange and eases the sidetone problem except for short subscriber-lines connected to resistive junction plant (e.g. PCM circuits);
- an impedance approximating the reference resistance (e.g. 600 ohms) eases standardization problems particularly in respect of alternative uses of the local line for non-speech services, but the optimum in respect of sidetone cannot be attained over the whole range of local line lengths.

References [2], [7] and [11] touch upon this subject.

8 Sidetone balance impedance

The degree of sidetone suppression is governed by the following parameters:

- microphone sensitivity;
- earphone sensitivity;
- sidetone balancing arrangement within the telephone instrument circuit;
- the impedance of the line to which the telephone is connected.

The microphone and earphone sensitivities and the instrument circuit are in part controlled by the required sending and receiving sensitivities. The impedance of the line to which the telephone is connected is not usually within the control of the telephone instrument designer. The only parameter freely available to the telephone designer to control the sidetone level is Z_{SO}, the sidetone balance impedance [7], [8], the impedance which when connected to the telephone completely suppresses sidetone (see also ref. [12]). If a transformer hybrid is used in the telephone then the internal balance network impedance is equal to the sidetone-balance impedance Z_{SO} modified by the turns ratio of the transformer. However, the concept Z_{SO} is not affected if the circuit uses any other form of balancing arrangement instead of a transformer.

9 Interworking with the existing network

The design of new handset telephones to be introduced into the telephone network must take account of the need to give satisfactory transmission on connections to existing local telephone circuits either directly or via the long-distance network. Reference [7] contains information touching upon this aspect.

Reference [10] is an example of a specification used in North America. Guidance for desirable sending and receiving levels are given as well as characteristics to be minimally acceptable for connection to the public switched network. It should be noted that this specification uses IEEE terminology, which is different from that found in CCITT Recommendations.

References

[1] CCITT Recommendation *Description of the ARAEN*, Green Book, Vol. V, Rec. P.41, Fig. 4, ITU, Geneva, 1972.

[2] CCITT Contribution COM XII-No. 32 (U.K. Post Office), Study Period 1973-1976.

[3] CCITT Contribution COM XII-No. 22 (Australia), Study Period 1973-1976.

[4] GLEISS (N.): Sound transmission quality, Tele. No. 1, 1972, pp. 44-53.

[5] CCITT Contribution COM XII-No. 229 (Sweden), Study Period 1985-1988.

[6] CCITT Recommendation *Subjective effects of direct crosstalk; thresholds of audibility and intelligibility*, Yellow Book, Vol. V, Rec. P.16, ITU, Geneva, 1981.

[7] CCITT *Manual Transmission planning of switched telephone networks*, Chapter V, Annex 1, ITU, Geneva, 1976.

[8] RICHARDS (D. L.): Telecommunications by speech, Chapter 5, *Butterworths*, London, 1973.

[9] CCITT Contribution COM XII-No. 105 (LME), Study Period 1973-1976.

[10] EIA Specification RS 470.

[11] CCITT Contribution COM XII-No. 144 (British Telecom), Study Period 1981-1984.

[12] CCITT Handbook on *Telephonometry*, Geneva, 1987.

Supplement No. 11

SOME EFFECTS OF SIDETONE

(Malaga-Torremolinos, 1984; amended Melbourne, 1988

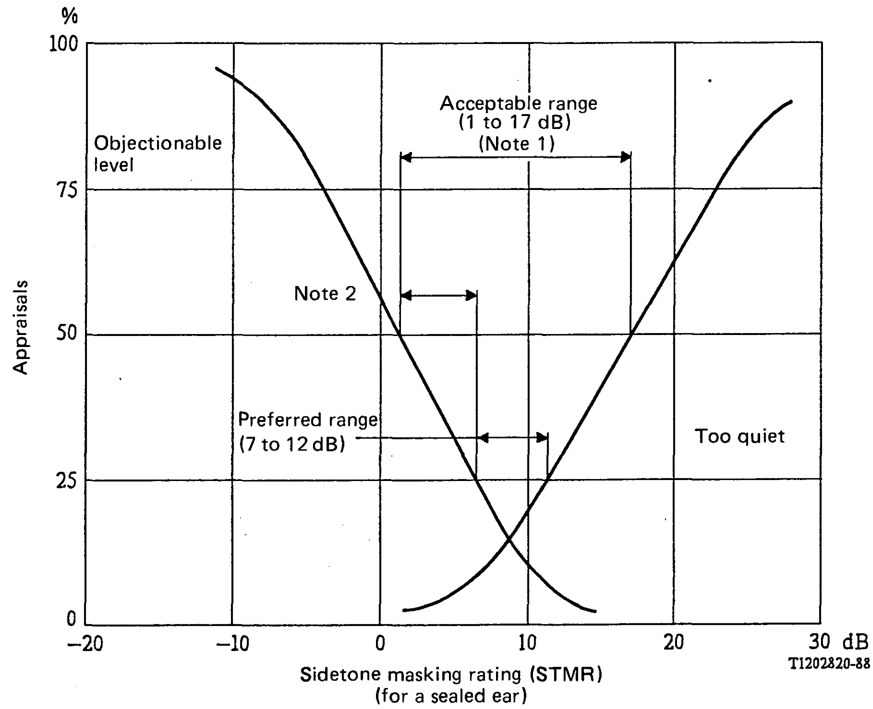
(referred to in Recommendations P.11 and P.79)

1 Introduction

Over a number of years sidetone has been studied in CCITT Study Group XII under Question 9/XII. Some important conclusions have been reached from the point of view of the subscriber in his role as both talker and listener. These conclusions relate to the effect of sidetone on a subscriber, as he hears his own voice, the way his talking level changes as a result and some effects of sidetone when the subscriber is listening in conditions of moderate to high-level room noise. These effects are summarized in Figures 1 and 3.

2 Talker sidetone

Figure 1 shows that there is a preferred range for sidetone when the subscriber is talking under quiet conditions, and that the difference between the sidetone being objectionable or too quiet is of the order of 20 dB. (These results were obtained from talking-only tests and need to be confirmed by conversation tests.) The preferred range lies between 7 and 12 dB, STMR (sidetone masking rating – Recommendation P.76) [1], [5].



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 to 7 dB) should only be entered with caution, e.g. on low loss connections, (see Recommendation G.121) or where there is a receive volume control.

FIGURE 1

Curves showing sidetone levels that are objectionable and too quiet, together with the preferred range, for the subscriber as a talker

The acceptable range is wider and lies between an STMR of 1 dB and 17 dB, (although it must be stated that increasing STMR to a value greater than 17 dB is likely to affect only the talking level, and that only marginally). This range corresponds to the difference between the two curves at the 50% appraisals level. It is not proposed that the 17 dB figure should in any way be considered a maximum value. However, for an STMR above 20 dB, the connection sounds "dead".

For telephone connections where the OLR is in the preferred range, the STMR values may similarly be positioned in the preferred STMR range given above. However, on high loss connections the STMR value should be close to, or even exceed 12 dB. On low loss connections the STMR value may be sometimes permitted to become less than 7 dB, but only rarely should it become as low as 1 dB, e.g. telephone sets with receive volume control. Recommendation G.121 interprets those results for transmission planning purposes.

Figure 2 shows the way in which the talking level changes with sidetone level [1], [2], [3], [4]. These results were obtained by means of conversation tests [6], for a connection close to the preferred overall loss. The speech voltage will also be a function of room noise for the same connection conditions.

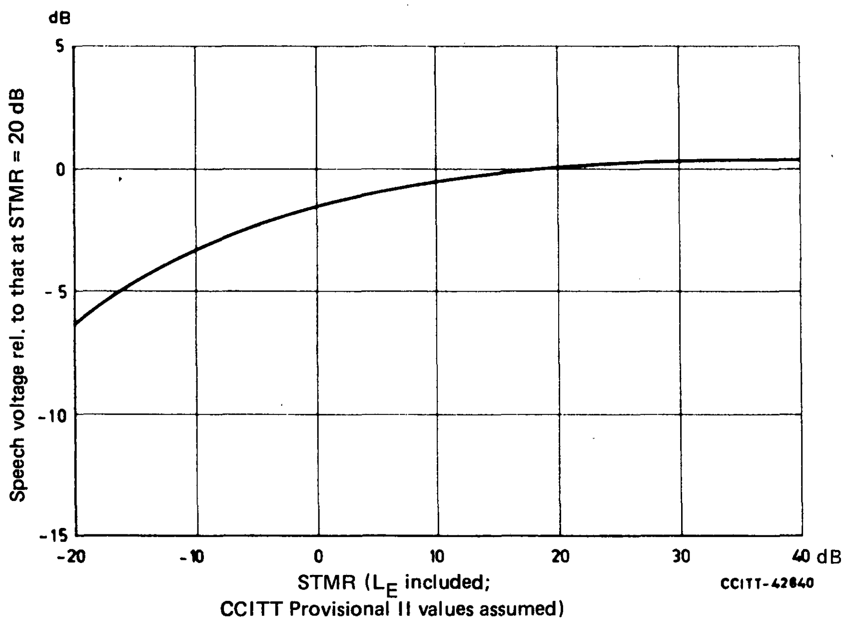


FIGURE 2

Speech voltage as a function of STMR

3 Listener sidetone

High room noise in the subscriber's environment disturbs the received speech in two ways:

i) noise being picked up by the handset microphone and transmitted to the handset receiver via the electric sidetone path,

ii) noise leaking past the earcap at the handset receiver.

Studies have shown that at low frequencies the earcap leakage path dominates over the electric sidetone path in much the same way as the human sidetone signal does in talker sidetone. The weightings applied in the STMR loudness calculation are therefore applicable and the listener sidetone rating (LSTR, Recommendation P.76) has been developed, which makes use of the room noise sidetone sensitivity (see Recommendation P.64, § 9) in the STMR rating method (Recommendation P.79).

Results of subjective tests from two Administrations [7], [8] (using in this case a mean opinion scale of 0-10) are given in Figure 3. In each case the LSTR was derived by making use of Δ_{Sm} (see Recommendations P.10, P.64, P.79 and the *Handbook on Telephonometry*, § 3.3.17c) to convert the sidetone sensitivities S_{mest} to S_{RNST} before calculating LSTR (Australian results) or applied as a weighted correction to STMR (Swedish results) as described in Recommendation G.111, § A.4.3.3. Room noise levels were comparable at 55-59 dBA.

Based upon these results Recommendation G.121 recommends that a value of 13 dB LSTR should be striven for.

The value 13 dB is based on a 10 dB LSTR (which may be considered a minimum value), where no further improvement in mean opinion score was possible by increasing LSTR (Figure 3), plus an allowance of 3 dB reflecting the fact that room noise in some office locations can exceed the values used in these experiments. Other tests (Sweden) have also suggested that a higher figure might be more appropriate.

The value that is satisfactory in a given telephone connection will depend on such factors as the level of room noise, the OLR of the connection, the talking levels used, etc. This is still under study in Question 9/XII.

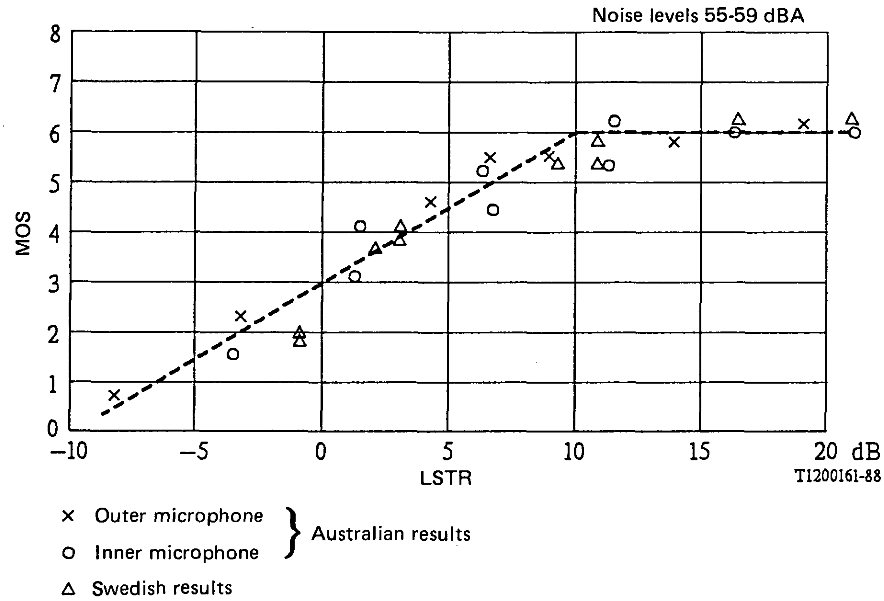


FIGURE 3

MOS as a function of LSTR calculated from Australian and Swedish test results

References

[1] CCITT Contribution COM XII-No. 50, Study Period 1977-1980 (ITT).

[2] CCITT Contribution COM XII-No. 171, Study Period 1977-1980.

[3] CCITT Contribution COM XII-No. 199, Study Period 1977-1980 (Australia).

[4] CCITT Contribution COM XII-No. 116, Study Period 1977-1980 (Hungary).

[5] CCITT Contribution COM XII-No. 152, Study Period 1981-1984 (NTT).

[6] Results of conversation tests sent directly to Special Rapporteur for Question 9/XII, British Telecom, 1978.

[7] CCITT Contribution COM XII-No.151, Study Period 1981-1984 (Australia).

[8] CCITT Contribution COM XII-No.70, Study Period 1985-1988 (Sweden).

Supplement No. 13

NOISE SPECTRA

(Malaga-Torremolinos, 1984)

(quoted in Recommendations P.44 and P.45 (Orange Book, Volume V) and Question 24/XII)

(Contribution from British Telecom)

1 Introduction

This Supplement gives the descriptions of noise spectra used in the evaluation of telephony transmission performance that are recommended by the CCITT or have been employed in studying questions assigned to Study Group XII.

Controlled environmental noise is used in subjective evaluations such as:

a) AEN determinations as described in Recommendations P.44 [1] and P.45 [2];

b) conversation and listening experiments as described, for example, in Supplement No. 2 [3].

Spectra for two different environments are described, one for room noise and two for internal vehicle noise.

2 Room noise

The room noise should have a power density spectrum corresponding to that published by Hoth [4]. Table 1 gives the spectrum density adjusted in level to produce a reading of 50 dBA on a sound level meter conforming to IEC Recommendation Publication 179 [5]. This is reproduced in Figure 1. This spectrum is independent of level, i.e. for 40 dBA the level in each band will be 10 dB less than that shown in Table 1. Additional information on the power in each 1/3rd octave band is also given in Table 1.

3 Interńal vehicle noise

Two spectra representing internal vehicle noise [6], [7] have been recommended for use in the study of Question 24/XII [8] for evaluating mobile radio systems. They are adequately represented by simplified curves [9]; one spectrum for moving vehicles and the other for stationary vehicles. Table 2 gives the spectrum densities together with additional information on the power in each 1/3rd octave band. The spectrum density for moving vehicles is shown in Figure 2 *a)* and for stationary vehicles in Figure 2 *b)*. These spectra are independent of level.

Table 3 gives the computed values of the unweighted sound pressure levels for various speeds calculated over the ISO 1/3rd octave frequency bands centred on 63 Hz to 8000 Hz.

TABLE 1

Room noise spectrum

Frequency (Hz)	Spectrum density (dB SPL/Hz)	Bandwidth 10 $\log_{10} \Delta f$ (dB)	Total power in each 1/3rd octave band (dB SPL)	Tolerance (dB)
100	32.4	13.5	45.9	± 3
125	30.9	14.7	45.5	
160	29.1	15.7	44.9	
200	27.6	16.5	44.1	
250	26.0	17.6	43.6	
315	24.4	18.7	43.1	
400	22.7	19.7	42.3	
500	21.1	20.6	41.7	
630	19.5	21.7	41.2	
800	17.8	22.7	40.4	
1000	16.2	23.5	39.7	
1250	14.6	24.7	39.3	
1600	12.9	25.7	38.7	
2000	11.3	26.5	37.8	
2500	9.6	27.6	37.2	
3150	7.8	28.7	36.5	
4000	5.4	29.7	34.8	
5000	2.6	30.6	33.2	
6300	−1.3	31.7	30.4	
8000	−6.6	32.7	26.0	

Note 1 – The electrical input signal, e.g. white noise, shall be band-limited to the 1/3rd octave bands centred on the ISO preferred frequencies (ISO 266) between 100 Hz and 8000 Hz with the band edges conforming to the filters described in IEC 225.

Note 2 – The acoustical room noise is difficult to control at low frequencies, especially in the unspecified region below 100 Hz because of the dimensions of typical test cabinets, poor attenuation of such cabinets and the influence of extraneous noises, e.g. air-conditioning plant. It is therefore desirable to select a test cabinet that keeps these unwanted low frequency sound pressure levels to a minimum.

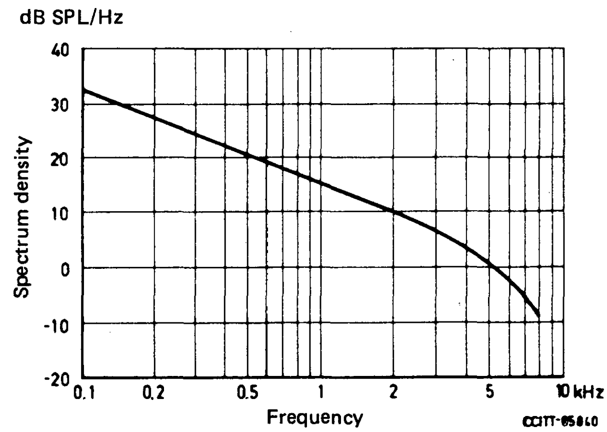
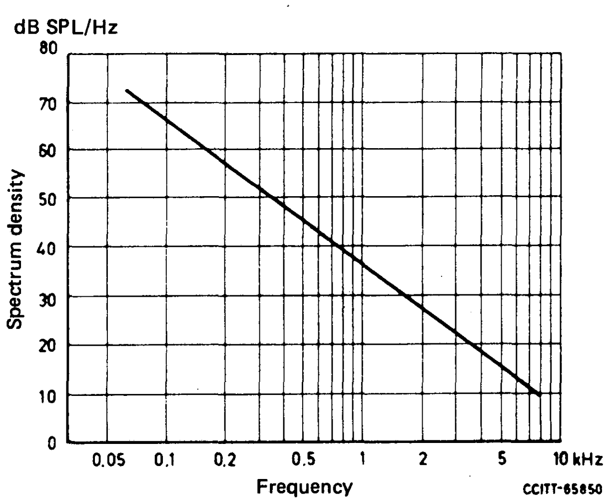
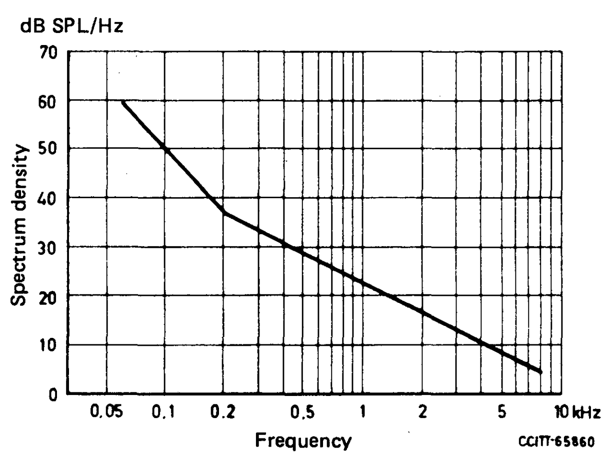


FIGURE 1

Room noise spectrum density



a) Spectrum density for moving vehicles



b) Spectrum density for stationary vehicles

FIGURE 2

TABLE 2

Internal vehicle noise spectra

Frequency (Hz)	Spectrum density (dB SPL/Hz)		Bandwidth $10 \log_{10} \Delta f$ (dB)	Total power in each ⅓ octave band (dB SPL)		Tolerance (dB)
	Moving	Stationary		Moving	Stationary	
63	72.3	58.3	11.7	84.0	70.0	± 3
80	69.3	55.0	12.7	82.0	66.7	
100	66.5	49.8	13.5	80.0	63.3	
125	63.3	45.1	14.7	78.0	60.0	
160	60.3	42.0	15.7	76.0	56.7	
200	57.5	36.8	16.5	74.0	53.3	
250	54.4	34.7	17.6	72.0	52.3	
315	51.3	32.6	18.7	70.0	51.3	
400	48.3	30.6	19.7	68.0	50.3	
500	45.4	28.7	20.6	66.0	49.3	
630	42.3	26.6	21.7	64.0	48.3	
800	39.3	24.6	22.7	62.0	47.3	
1000	36.5	22.8	23.5	60.0	46.3	
1250	33.3	20.6	24.7	58.0	45.3	
1600	30.3	18.6	25.7	56.0	44.3	
2000	27.5	16.8	26.5	54.0	43.3	
2500	24.4	14.7	27.6	52.0	42.3	
3150	21.3	12.6	28.7	50.0	41.3	
4000	18.3	10.6	29.7	48.0	40.3	
5000	15.4	8.7	30.6	46.0	39.3	
6300	12.3	6.6	31.7	44.0	38.3	
8000	9.3	4.6	32.7	42.0	37.3	

TABLE 3

Computed sound pressure levels of spectra

Spectra		Sound pressure level, unweighted (dB SPL)
Moving	30 km/h	80
	80 km/h	85
	110 km/h	90
Stationary		75

Notes to Tables 2 and 3:

Note 1 – These values apply for typical vehicles. Discretion may be used to adjust the levels downwards for luxury vehicles and upwards for noisier vehicles.

Note 2 – Because of the practical difficulty of generating such high sound pressure levels at low frequencies, and because normal speech contains no apparent energy below about 63 Hz in which range of frequencies the car is also comparatively intensitive it is probably advisable to restrict the recommended noise spectrum to frequencies above 63 Hz. However, it should be borne in mind that low and medium frequency vibrations have important physiological and psychological effects which should be sudied in their own right.

Note 3 – The electrical input signal, e.g. white noise, shall be band-limited to the 1/3rd octave bands centred on the ISO preferred frequencies (ISO 266) between 63 Hz and 80000 Hz with the band edges conforming to the filters described in IEC 225.

Note 4 – The acoustical room noise is difficult to control at low frequencies especially in the unspecified region below 63 Hz because of the dimensions of typical test cabinets, poor attenuation of such cabinets and the influence of extraneous noises e.g. air-conditioning plant. It is therefore desirable to select a test cabinet that keeps these unwanted low frequency sound pressure levels to a minimum.

References

[1] CCITT Recommendation *Description and adjustment of the reference system for the determination of AEN (SRAEN)*, Yellow Book, Vol. V, Rec. P.44, ITU, Geneva, 1981.

[2] CCITT Recommendation *Measurement of the AEN value of a commercial telephone system (sending and receiving) by comparison with the SRAEN*, Yellow Book, Vol. V, Rec. P.45, ITU, Geneva, 1981.

[3] *Methods used for assessing telephony transmission performance*, Supplement No. 2, Yellow Book, Vol. V, ITU, Geneva, 1981.

[4] HOTH (D. F.): Room noise spectra at subscribers' telephone locations, *J.A.S.A.*, Vol. 12, pp. 499-504, April 1941.

[5] IEC Recommendation Publication 179, *Precision sound level meters*, 1965.

[6] CCITT Question 24/XII, Contribution COM XII-No. 120, (Noise inside light motor vehicles), Study Period 1981-1984.

[7] CCITT Question 24/XII, Contribution COM XII-No. 134, (Internal vehicle noise spectra), Study Period 1981-1984.

[8] CCITT Question 24/XII, Contribution COM XII-No. 1, (Link with mobile stations), Study Period 1981-1984.

[9] CCITT Contribution COM XII-No. 208, (Comparison of the results of vehicle noise submitted by France and BT), Study Period 1981-1984.

Supplement No. 14

SUBJECTIVE PERFORMANCE ASSESSMENT OF DIGITAL PROCESSES USING THE MODULATED NOISE REFERENCE UNIT (MNRU)

(Malaga-Torremolinos, 1984; amended Melbourne, 1988)

(quoted in Recommendation P.81)

1 Introduction

The primary purpose of this Supplement is to define a specific subjective testing method for evaluating digital processes in a manner such that the quantization distortion effects of these processes on transmission performance can be taken into account in the evolving international telephone network. This implies both the ability to uniquely assign a numerical contribution to each digital process and the ability to use this assigned contribution in conjunction with other impairments to estimate telephone connection performance.

Secondary purposes of the Supplement are to suggest ways in which the subjective test results can be treated to arrive at the assigned impairment level of a particular digital process and how this assigned impairment level can be used in transmission performance analysis.

2 Impairment reference scale for digital processes

Two reference scales that have been used for performance assessment of digital processes are a) continuous random noise (additive noise) and b) random noise with amplitude proportional to the instantaneous signal amplitude (multiplicative noise). Random noise with amplitude proportional to the instantaneous signal amplitude in terms of the Q ratio, according to the MNRU as specified in Recommendation P.81, should be used.

The reasons for this proposal are:

1) The signal processed through the MNRU is perceptually very similar in character to the signal processed through various digital processes, thus resulting, in principle, in easier assessment by test subjects, and

2) Considerable experience and information have been accumulated with the MNRU.

Note – It has not been documented that Q represents a more suitable reference scale than continuous random noise.

3 Survey of methods

A number of methods are suitable for characterizing the performance of digital processes in terms of Q values. The methods deal with in this Supplement comprise listening-only tests. They are summarized in Table 1.

Other possible methods that may be mentioned are:

1) multiple paired comparisons between all systems under test and all reference condition $/X_1/X_2 \dots X_i/X_j \dots X_j/R_i$,
2) articulation test of MNRU conditions and digital systems in the same experiment.

These methods are not described here.

TABLE 1

<table>
<tr><td colspan="4">Indirect comparison with MNRU</td><td colspan="2">Direct comparison with MNRU</td></tr>
<tr><td colspan="2">ACT method</td><td colspan="2">DCR method</td><td colspan="2">Equality Threshold</td></tr>
<tr><td>X_1
X_2
.
.
.
X_n
R_i</td><td>(SSR)</td><td>X_0/X_1
X_0/X_1
.
.
.
X_0/X_n
X_0/R_i</td><td>(PC)</td><td>X_1/R_i
X_2/R_i
.
.
.
X_n/R_i</td><td>(PC)</td></tr>
<tr><td colspan="2">Annex A</td><td colspan="2">Annex B</td><td colspan="2">Annex C</td></tr>
</table>

$X_1, X_2, \dots X_n$	Digital systems under test
X_0	Unprocessed (reference) system
R_i	MNRU reference (Q level "i")
PC	Paired Comparisons
SSR	Single Stimulus Ratings
ACR	Absolute Category Rating
DCR	Degradation Category Rating

4 Background for the test methods

4.1 *Background for the Absolute Category Rating (ACR) test method of Annex A*

The method is based on a procedure utilized in an experiment conducted by a working group of the IEEE (Institute of Electrical and Electronics Engineers) in which representatives from seven countries participated (Canada, France, Italy, Japan, Norway, the United Kingdom and the United States) [2]. The aim of this experiment was to determine whether comparable results could be obtained when the same test is performed in several countries. Speech samples in the native languages of the participating countries were processed at a central location through 38 communications circuits. The recordings of the processed speech were returned to each country for evaluation on a five-point category rating scale by native listeners.

The communications circuits included 22 references (continuous random noise, MNRU, μ-255 PCM) and 16 adaptive differential PCM (ADPCM) systems. (The type of ADPCM system used was a first order fixed predictor [3].) An important part of the data analysis was estimation of the quality of the 16 ADPCM conditions at one location given measurements of ADPCM quality elsewhere.

Results (mean opinion scores [MOS]) obtained at the different locations differed [2]. Nonetheless, analysis of the results indicated that a reasonably accurate estimate of ADPCM quality in country B is the quality measured in country A adjusted by an additive constant.

Changes in the methodology were discussed at an IEEE working group meeting in May 1982 in Paris. The methodology incorporating these changes was recommended by Study Group XII in June 1982 as a basis for evaluating candidate 32 kbit/s algorithms for CCITT standardization as Recommendation G.721 [1]. Subjective tests using the methodology were conducted under the auspices of Study Group XVIII in late 1982 with the results that a codec algorithm was selected and improvements (not related to telephone speech transmission issues) were identified. A second series of subjective tests in late 1983 confirmed that telephone speech transmission performance of the improved algorithm was suitable. (Differences between test results from the different participating organizations were also found in the 1982 and 1983 CCITT tests.)

The preceding discussion should not be taken to indicate that the subjective testing methodology is completely satisfactory: the reasons for differences found between countries [2] and [4] are thus far not explained. Nevertheless the testing methodology has the important feature of having been used by several countries.

4.2 *Background for the Degradation Category Rating (DCR) test method of Annex B*

A modification of the ACR test method, the DCR test method, is described in Annex B. Based on results from one Administration, the DCR test method provides a greater discrimination between conditions than does the ACR test method [5].

Results from an experiment conducted by another Administration do not support this conclusion [6].

5 Analysis of test results

The purpose in conducting test of digital processes is to determine their suitability for use in telephone networks. A procedure which has been used is to assign Q values, determined using the reference system of Recommendation P.81, to processes of interest. Various methods of data analysis are possible, but it appears desirable to define a single method to be used in order to assure expressing results in common terms. The provisional method is based on the use of MOS (mean opinion score) values obtained using the procedures of Annex A.

Hypothetical results obtained from a subjective test conducted according to the methodology of Annex A to this Supplement are shown in Figures 1 to 4. (Straight lines are used simply to connect data points.) Generally such results will display a saturation effect at and near the very good conditions (high MOS) and the very bad conditions (low MOS). (For high MOS, the saturation is caused by the 5-point scale and possibly by the idle circuit noise of the subjective test system without added impairments, e.g. idle circuit noise, and codec quantization distortion. For low MOS, the saturation is caused by the 5-point scale.) Experience [2] has shown that due to this saturation effect, acceptable accuracy for the determination of Q is obtained for the range of about 5 dB to 25 dB.

An objective of this analysis is to determine a function $Q_2 = F(L)$ where Q_2 is the Q value for the code and L is the line bit rate. One simple method for determining this function uses the MOS values shown in Figures 2 and 3 and can produce a graph of this function as shown in Figure 5. The method is shown in Figure 6, wherein a value of line bit rate is chosen, say L_2, and its corresponding MOS value is determined. This MOS value is then used to enter the right hand graph to find the value of Q, in this case Q_2, corresponding to this MOS value. Q values for all the other L values are obtained in a similar way and the resulting set of (L_i, Q_i) gains are plotted as in Figure 5.

Analysis of test results should include statistical analysis to establish that MOS values obtained are due to the test conditions and not to other factors. Student's test may be suitable, but there is some indication that analysis of variance is more appropriate.

The principles of a method of analysis used by one organization are outlined in Annex D of this Supplement. The method uses analytic values, called fit means, calculated from subjective test results; these analytic values are similar to MOS values calculated from test results. One desirable result of the test is estimates of the Q of the processes tested. Annex D contains a method for deriving such estimates.

Values of MOS versus Q (as per Figure 2) obtained from actual experiments are given in References [5], [7] and [18] and in Annex B.

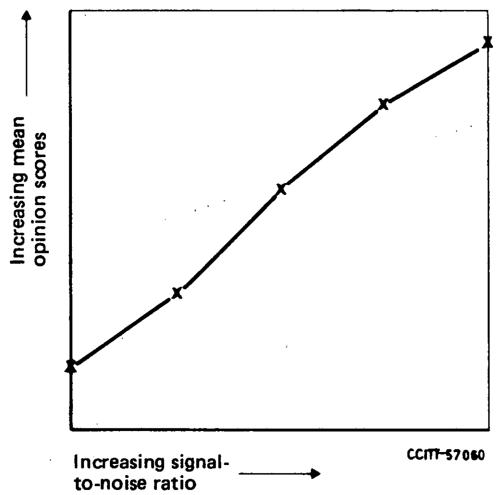


FIGURE 1

Mean opinion scores for SNR; speech power in dBm, idle circuit noise power in dBmp

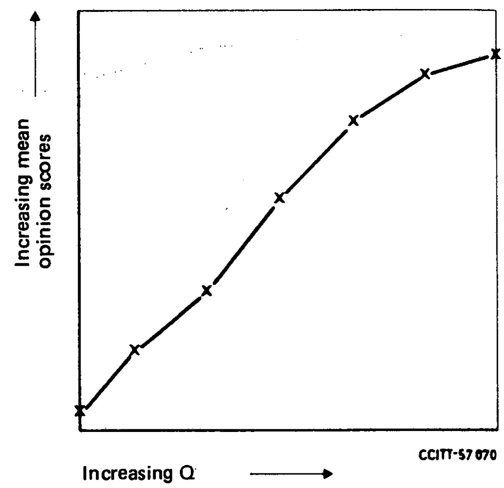


FIGURE 2

Mean opinion scores for Q; Q is the setting of the MNRU

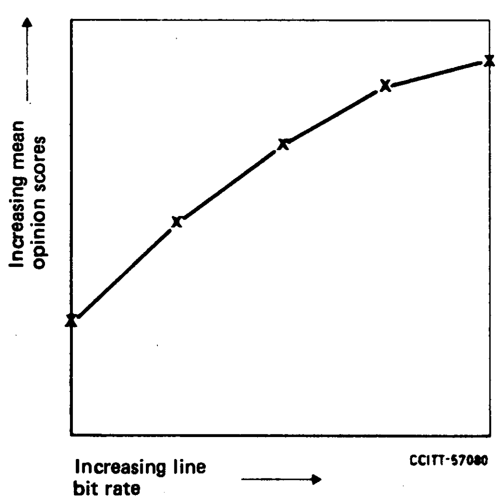


FIGURE 3

Mean opinion scores for a single codec at different line bit rates

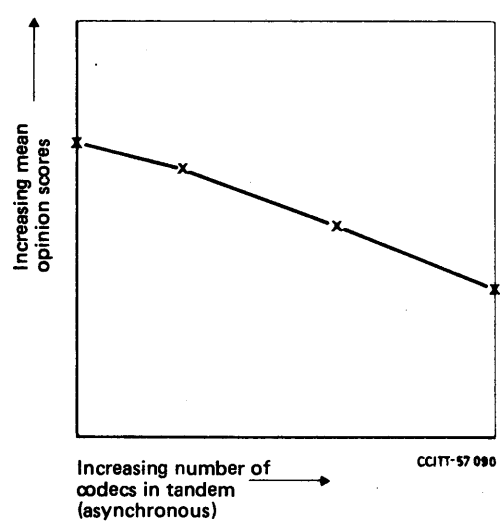


FIGURE 4

Mean opinion scores for tandem codecs

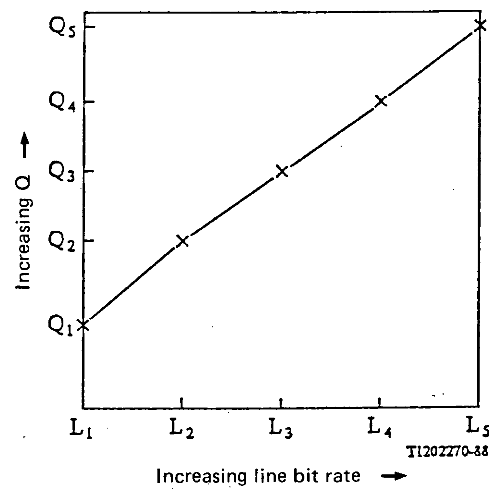


FIGURE 5

Q as a function of line bit rate for a single codec

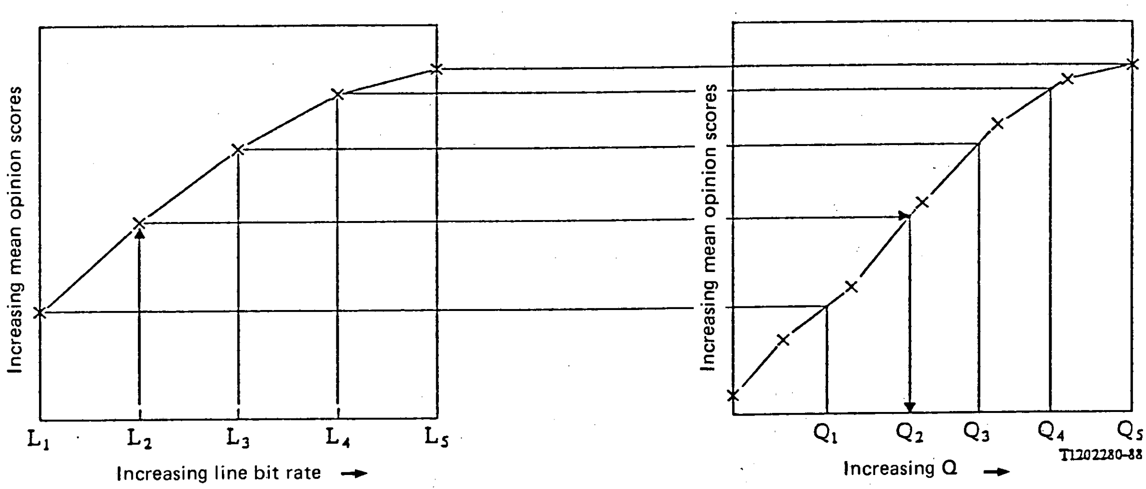


FIGURE 6

Graphical method for deriving Figure 5 from existing Figures 2 and 3

ANNEX A

(to Supplement No. 14)

**Absolute category rating (ACR)
method for subjective testing of digital processes**

A.1 *Introduction*

The listening-only test method consists in principle of three parts: preparation of source tapes; processing of the source tapes to obtain stimulus tapes containing the test conditions of interest; conduct of subjective tests using the stimulus tapes. Certain steps may be combined if interchange of source/stimulus tapes between locations is not involved.

The methodology is based on the notion of simulating a connection comprising a sending system, a receiving system and an interconnection system which provides for inserting the impairment of interest (idle channel noise and quantization distortion from the MNRU and from digital processes).

Listener responses in the subjective tests are influenced by a number of sources of variation, e.g. speech material, talker voice characteristics, presentation orders, time effects, etc. Unless controlled in some way, these variables may bias the outcome of the experiment. It is therefore recommended that appropriate experimental design be applied to take this into account. Principles for experimental design may be found in textbooks on statistics.

A.2 *Preparation of source tape(s)*

The recording system consists of a tape recorder, means for injecting calibration tones and a suitably defined sending system.

A.2.1 *Tape recorder*

The tape recorder should be a high (studio) quality two-track machine. The type of equalization should be stated, but IEC is preferred. One of the tracks is used for recording the speech samples; the second channel is available for other purposes, e.g. cueing tones to allow computer start/stop control of the tape recorder. The tape recorder should be operated at 19 cm/sec.

Low print-through, low-noise tape should be used and the tape should be stored "tail-out" so that it is necessary to rewind the tape before it is played.

Note – The use of an A/D converter and a television cassette recorder should be considered as a means for recording and storing high quality source and test tapes.

A.2.2 *Calibration tones*

It is recommended that calibration tones be recorded on the source tape(s) to enable checking the sensitivity/frequency characteristics of the connection simulation from input to the source tape recorder to output from the stimulus presentation tape recorder. Tones should be recorded in sequence at 250, 500, 1000, 2000, 3000 and 4000 Hz of 5 seconds duration each, with a level 6 dB below the maximum r.m.s. input level of the tape recorder. These tones should be followed by a 15 second recording of a 1 kHz test tone at maximum r.m.s. input level to enable calibration of the interconnection and listening systems. This should be followed by several metres of leader tape.

A.2.3 *Sending system*

The sensitivity/frequency characteristics of sending systems of different countries are likely to differ and, thus, results of different countries may differ because of attenuation distortion. Furthermore, the performance of complex digital codec algorithms may be dependent on the shape of the sending system sensitivity/frequency characteristics. Therefore, it is desirable that at least for some of the conditions in a test the sending system characteristic be as given in Table A-1 (simulates the IRS send part without filter).

TABLE A-1

IRS characteristics before adding SRAEN filter

Frequency (Hz)	S_{MJ} (dB V/Pa)	S_{je} (dB Pa/V)
100	−22.00	−21.00
125	−18.00	−17.00
160	−14.00	−13.00
200	−10.00	− 9.00
250	− 6.80	− 5.70
315	− 4.60	− 2.90
400	− 3.30	− 1.30
500	− 2.60	− 0.60
630	− 2.20	− 0.10
800	− 1.20	+ 0.00
1000	+ 0.00	+ 0.00
1250	+ 1.20	+ 0.20
1600	+ 2.80	+ 0.40
2000	+ 3.20	+ 0.40
2500	+ 4.00	− 0.30
3150	+ 4.30	− 0.50
4000	+ 0.00	−11.00
5000	− 6.00	−23.00
6300	−12.00	−35.00
8000	−18.00	−53.00

It may be desirable to include conditions for which the sending system represents a typical (average) local system according to the testing organization's (country's) network and/or needs. This system comprises a handset telephone set, a simulated physical cable pair, a feeding bridge and a resistive termination (e.g. 600 ohms, 900 ohms) to which the source tape recorder is connected. The telephone set can utilize a linear telephone microphone with a real voice sensitivity/frequency characteristic such that the acoustic-to-electric response of the sending system represents the organization's average local system. It may also be desirable to include conditions obtained with a carbon telephone microphone representative of the type(s) used in the organization's (country's) network. (See Recommendation P.64.) The characteristics (and feeding current) should be reported. It may also be desirable to report the characteristic measured using an artificial sound source. (See Recommendations P.51 and P.64.)

A.2.4 *Recording environment*

The recording environment should be that of a quiet living room or office. The ambient room noise level should be 25-30 dBA. The noise spectrum should, if possible, have the shape of the Hoth spectrum of Supplement No. 13. Special tests may be required using other noise levels and/or spectral characteristics (e.g. typewriter noise, etc.).

The room noise characteristic should be reported in as complete a form as is possible [e.g. dBA, long-term spectrum, amplitude/time distribution, etc.].

A 30 second recording of the room noise through the local system should follow the calibration tones. This should be accomplished with a talker holding the telphone handset in a normal use manner. (Special precautions may be necessary in order to avoid breath sounds if desired.)

A.2.5 *Speech samples*

A source tape is made of 4 × C samples (4 talkers, samples consisting of training, reference and test conditions). Each sample should comprise 2 or 3 sentences separated by at least 1 second.

All samples should be different to avoid repetition of sentences during a test. When reporting test results, it may be desirable to provide a list of the sentences used (i.e. 8 × C or 12 × C sentences).

Each sample is expected to be 6-10 seconds in length. The samples should be separated by 5 seconds of silence to allow for control (e.g. turning the tape recorder on and off) and of the amount of time needed for subjects to vote.

The r.m.s. level of the speech samples (speech power while active) should be 12 dB below the r.m.s. level of the 1 kHz calibration tone in order to avoid peak clipping of the speech samples by the tape recorder and to measure in an easy way the actual r.m.s. level of the speech.

A.2.6 *Talkers*

At least 4 different talkers (2 female, 2 male) with different voice characteristics should be used. Selection of the talkers will depend on the judgement of the experimenter.

A.3 *Preparation of stimulus tape(s)*

The interconnection system will consist of the source tape recorder (resistive, 600 or 900 ohms), an input filter, a means for inserting test conditions, an output filter, and the stimulus tape recorder (resistive, 600 or 900 ohms). The characteristics of the filters should be provided.

A.3.1 *Test conditions*

The test conditions comprise the digital codec(s) of interest. The codec(s) should be defined as simply and completely as possible (e.g. A-law/μ-law, ADPCM with first order fixed predictor, etc.). This is to enable unique performance specification for codecs of the same type.

Because codecs may have different performances at different speech input levels, they should preferably be tested not only at a nominal fully-loaded condition, but also at levels below and above this level, say ± 10 dB. These changes in input level to codecs should be "off-set" by corresponding adjustments of their outputs to maintain an approximately constant output level for the test. (Listening level may also affect relative performance of different digital processes. See also § A.4.4.)

The codec(s) should be tested singly (one encoding/decoding pair) and with 2, 4 and (possibly) 8 codecs connected in tandem asynchronously. (It may also be desirable to include conditions in tandem synchronously.) The codecs may be hardware or software implemented; if the latter, injected circuit noise expected for practical codecs should be included.

For the single codec(s) conditions, the line bit rate should be the design value and, if possible, line bit rates both above (to ensure subjective saturation) and below (to ensure degraded performance). These conditions may be useful in assigning a performance level(s) to the codec(s). (For example, a nominal 32 kbit/s ADPCM algorithm might also be tested at 16, 24, 40 and 48 kbit/s.)

The tandem conditions should utilize the codec(s) at the design line bit rate(s).

Codec conditions with line errors should be included. Bit error rates covering the range 10^{-3} to 10^{-6} should be used.

A.3.2 *Reference conditions*

Reference conditions which should be included are Q values within the range 5 dB to 25 dB with a minimum of 4 steps. (It may also be desirable to include Q values of 0 dB and 30 dB.)

It is desirable to include injected circuit noise values to provide SNRs within the range 5 dB to 45 dB with a minimum of 4 steps. (SNR is the dB ratio of speech power in milliwatts while active to injected circuit noise in milliwatts; the circuit noise conditions should be band-limited by filters having the same characteristics as the filter of the MNRU.) Note that the 45 dB ratio could be dependent on the inherent system noise, e.g. noise from the source tape preparation process, noise from the source tape recorder, etc.

Source conditions should also be included. (These are obtained by removing the injected idle circuit noise.)

The purpose of including the injected circuit noise conditions is to enable the relating of test results to results available on the effects of loss and circuit noise (Question 4/XII) and to allow use of the test results in subjective opinion model studies (Question 7/XII).

Other reference conditions can be included at the discretion of the testing organization. For example, particular organizations may have available information from previous tests of A-law/μ-law companded PCM, and it may be desirable to include some PCM conditions to allow comparison with previous results.

A.3.3 *Calibration*

The insertion loss of the interconnection circuit should be 0 dB at 1 kHz between the resistive source/termination. This should apply for the better conditions e.g. Q = 25 dB, SNR = 45 dB and the test codec(s) operated at design line bit rate(s).

The r.m.s. level of the 1 kHz calibration tone at the input to the inter-connection circuit should be 3 dB below the codec(s) overload level (which should be quoted). This will ensure that r.m.s. level of the speech samples will be 15 dB below the codec(s) r.m.s. sinewave overload level.

With the above calibration, the injected circuit noise levels in dBm across the output resistive termination should be adjusted to an appropriate level relative to the output 1 kHz calibration tone level in dBm. Note that in particular the circuit noise impairment should be present during the speech sample idle periods but not before and after the speech sample.

The stimulus tape recorder calibration should be the same as that for the source tape recorder.

A.3.4 *Stimulus tape(s)*

Stimulus tapes should begin with the 1 kHz calibration tone recorded (without introduced impairments), 12 practice conditions and then the test and reference conditions.

The practice conditions should be selected to introduce the test subjects to the test format and range of speech quality. These conditions should consist of each of the four talkers with 3 practice conditions.

The basic test and reference conditions will be 4 (i.e. number of talkers) times the number of nominal conditions. These conditions should appear in random order. There should be at least 2 stimulus tapes with different random orders. (These could be used in different tests with different subject groups.)

It may also be desirable to include replication of at least some of the test/reference conditions. However, this may not be possible for a practical subjective test size.

The timing of conditions in the stimulus tapes is the same as that for the source tapes, e.g. approximately 6-10 seconds (2 or 3 sentences) with each condition separated by 5 seconds of silence.

The calibration tones on the source tape need not appear on the stimulus tape (except for 1 kHz calibration tone as noted above). However, the calibration tone levels should be measured at the interconnection system output resistive termination so that the system sensitivity/frequency characteristics can be measured and reported for all condition types.

A.4 *Testing procedure*

A.4.1 *Listeners*

The preferred number of listeners is 32, assigned equally to each tape. At least 12 test subjects should be used. It is desirable that the subjects be selected to represent the typical customer population (e.g. half of the group females and half males, ages approximating the population distribution of ages, normal hearing, etc.).

A.4.2 *Listening system*

For reasons given in the first paragraph of § A.2.3, the receiving system characteristic should be as given in Table A-1 (simulates the IRS receive part without filter).

It would be desirable if the listening system simulated the organization's typical (e.g. average) local system representing the central office source impedance, feeding bridge, physical cable pair and the handset telephone set. The electric-to-acoustic sensitivity/frequency characteristic of the listening system should be determined (see Recommendation P.64). Sidetone in the listening system should be suppressed.

A.4.3 *Listening environment*

The listening handset(s) should be located in a room with an ambient room noise level $\leqslant$ 40 dBA, preferably 25-30 dBA (simulating a quiet office or living room). The noise spectrum should, if possible, have the shape of the Hoth spectrum of Supplement No. 13. The actual ambient room noise level and spectrum, if different from the above, should be reported.

A.4.4 *Speech level*

The 1 kHz calibration tone on the stimulus tape when played through the listening system should be adjusted such that reproduction occurs at a level of -3 dBPa as measured with the artificial ear recommended by the CCITT. (See Recommendation P.51.) This will result in a speech level of about -15 dBPa which is close to the preferred level. It may also be desirable to include conditions with a 10 dB lower level and 10 dB higher level since the listening level may affect the relative performance of different digital processes. [1)]

A.4.5 *Test instructions*

Test subjects will be provided with a written set of instructions which will also be read to them (either by the test administrator or by means of a tape recording). The instructions should be given before the practice conditions. Subjects should not be instructed that the practice conditions represent the full range of quality to be encountered in the test. After the practice conditions, there should be sufficient time allowed for answering possible questions by the subjects.

The subjects should be instructed to rate the conditions according to the five point quality scale as follows:

Score	*Quality rating*
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

In countries for which English is not the native language, the appropriate terms in the native language should be used.

Before the listening test is conducted, it is necessary to carry out practice sessions to ensure full adaptation of listeners to the test conditions and obtain a stable evaluation.

[1)] There is some indication that a speech level of -5 dBPa (1 kHz tone level of $+7$ dBPa) would be more suitable than -15 dBPa for discrimination between coder conditions.

A.4.6 *Data collection*

Subjects' responses can be recorded by computer, on paper or by such other means as are appropriate. If paper and pencil are used, the response to each condition should be recorded on a separate card so that the subject is not looking at a previous opinion while making a new judgement.

A.5 *Results reports*

Reporting all of the raw data may be desirable but results in excessive documentation. Therefore, it may be appropriate to combine data across talkers and report the number of ratings in each of the 5 categories for each condition type, e.g. Q = 15 dB, SNR = 25 dB, etc. (Conclusions resulting from an analysis of the study of possible talker effects should be reported.) In addition, mean opinion scores (MOSs), standard deviation, 95 percent confidence intervals and other statistics computed by the organizations in analyzing the data should be reported.

Other items which should be reported are as follows:

a) microphone type;

b) sensitivity/frequency characteristic of the sending system (Recommendation P.64);

c) description of recording room and ambient noise levels;

d) measurement and adjustment procedure for speech levels;

e) sensitivity/frequency characteristics of the interconnection system for all test/reference condition types;

f) sensitivity/frequency characteristic of the listening system (Recommendation P.64);

g) description of the listening room and ambient noise level;

h) method of recording test subject opinions;

i) description of subject group including age, sex, population, prior experience and, if possible, audiometric threshold;

j) handset dimensions.

Bibliography for Annex A

KIRK (R. E.): Experimental design procedures for the behavioral sciences, *Brooks/Cole Publishing Company*, Belmont California, 1968.

CCITT Recommendation P.64.

CCITT Recommendation P.74.

ANNEX B

(to Supplement No. 14)

Subjective perfomance assessment of digital encoders using the degradation category rating procedure (DCR)

(Contribution of the French Administration)

B.1 *Introduction*

A listening-only test method has been drafted by CCITT SG XII to assess the subjective quality of digital encoders (see Annex A). This procedure, Absolute Category Rating test (ACR), leads to a low sensitivity in distinguishing among good telephone quality coders (within the range of quality of 6-8 bit PCM coders). If higher sensitivity is needed we propose to use a modified version of that procedure, which can be defined as a Degradation Category Rating test (DCR). For image testing CCIR [6] recommends two alternative methods, absolute category ratings and degradation category ratings. The DCR procedure, which in particular uses an annoyance scale and a high quality reference before each judgement, seems to be suitable for evaluating good quality images. Therefore this method has been adapted to evaluate speech quality.

This Supplement first describes the adaptation of the DCR procedure to speech. Then the sensitivity of the method is compared with that of the ACR procedure on the same circuits. Only the differences between ACR and DCR procedure are presented here. One can refer to Annex A for common points which are not covered in this Annex.

B.2 *Degradation category rating procedure (DCR)*

B.2.1 *Speech samples*

Each configuration is evaluated by means of judgements upon four talkers reading two different samples. Each sample should comprise two sentences separated by at least one second. These two samples (S1, S2), hence four different sentences, should be selected from a wider corpus composed of phonetically balanced sentences so that the mean score obtained in evaluating MNRU circuits for these four sentences is about the same as that obtained for the wider corpus. Therefore the corpus consists of eight samples defined as follows:

talker T1 reading samples S1, S2

talker T2 reading samples S1, S2

talker T3 reading samples S1, S2

talker T4 reading samples S1, S2.

This results in a repetition of the two samples during the test. But we feel that this is not so critical for the procedure where a degradation is evaluated with regard to a reference, especially for good telephone quality where the intelligibility of speech is nearly perfect. The use of different samples for each configuration as is done in ACR experiments could be one of the reasons for this procedure's lack of sensitivity.

B.2.2 *Reference conditions*

Reference conditions should include multiplicative noise with Q values within the range of 10 to 30 dB with a minimum of four steps. (It may also be desirable to include Q values of 5 dB and 35 dB).

A high quality reference should be chosen to be inserted before each judgement. Usually source conditions are used, i.e. samples with no more degradation than those introduced by sending systems and limitations of frequency bandwidth. Four "null pairs" (A-A) are included to check the quality of anchoring of the listeners' judgements.

B.2.3 *Stimulus presentation*

The stimuli are presented to listeners by pairs (A-B) or repeated pairs (A-B-A-B) where A is the high quality reference sample and B the same sample processed by a codec. The purpose of the reference sample is to anchor each judgement of the listeners. Using a reference and subjective judgements with respect to that reference is quite a common procedure in psychoacoustics. It tends to result in a good sensitivity for the overall evaluation by listeners. Samples A and B should be separated by 0.75 s and in a repeated pair procedure (A-B-A-B) the separation between the two pairs should be 2 s.

It seems that the classical order effect observed in a one-sample listening test (ACR for example) is not observed with the DCR procedure. Thus, only one random order of presentation can be used. Therefore the basic test and reference conditions will be eight times (four talkers × two samples) the number of nominal conditions.

The timing for the response of listeners is the same as for the ACR test, i.e. 5 s between each presentation (pair or repeated pairs).

B.2.4 *Test instructions*

The subjects should be instructed to rate the conditions according to the five point degradation category scale as follows:

5 – Degradation is inaudible

4 – Degradation is audible but not annoying

3 – Degradation is slightly annoying

2 – Degradation is annoying

1 – Degradation is very annoying.

B.3 *Comparison between the sensitivity of an ACR and a DCR procedure for the same coder configurations*

Tables B-1 and B-2 summarize the results obtained with ACR test and DCR test respectively for the evaluation of three 32 kbit/s ADPCM algorithms.

Figures B-1, B-2 and B-3 show the mean opinion score (MOS) and degradation mean opinion store (DMOS) obtained by the same conditions with the two procedures (ACR and DCR respectively).

From these figures one can note:

- a good agreement between the results obtained with the two procedures;
- a larger spread of the DMOS obtained for MNRU circuits with Q values ranging from 10 dB to 35 dB, and a good anchoring of the judgements of listeners ("null pairs" have obtained a score of 4.98);
- a higher sensitivity of the DCR procedure in the range of good telephone quality ($20 < Q < 35$ dB).

These sensitivities can be quantified by means of a statistical multiple comparison test. When an *a posteriori* comparison of codecs is needed a Tuckey [7] honestly significant difference (HSD) test can be applied effectively. The HSD test is designed to make all pairwise comparisons among the means and to determine the significance of the differences in the mean values. Under identical conditions ($\alpha = 0.01$, $k = 2$, $N = 225$, fixed mode) the HSD limit value ($q_{\alpha,k,N}$) is 3.70 and since the residual errors for ACR and DCR procedures are about the same (0.42), two means can be declared as significantly different if:

$$\Delta = | \overline{X_i} - \overline{X_j} | > 0.21$$

This difference, expressed in Q value, corresponds to:

Range in Q (dB)	ACR test Δ	Range in Q (dB)	DCR test Δ
15 – 20	1.48	15 – 20	1.07
20 – 25	1.87	20 – 25	1.14
25 – 30	3.00	25 – 30	1.36

This means that the resolution of the DCR test may be twice that of the ACR test in terms of Q value in the range of good telephone quality.

B.4 *Conclusion*

A good agreement between the results obtained with the two procedures (ACR and DCR) has been found. The presence of a reference before each judgement for the DCR procedure ensures a good anchoring of the listener's rating and consequently a larger spread of the degradation mean opinion score (DMOS) obtained by the coders. The evaluation of the coders based on the same speech samples leads to a better precision for the DCR procedure at a price, of course, of a decrease of the importance of the effort made to comprehend the samples in the overall quality judgement. Therefore the degradation category rating procedure seems well adapted to evaluate good telephone quality coders.

TABLE B-1

Mean opinion scores (MOS) and 95% confidence intervals (INT) for ACR test

Test conditions	X		Y		Z	
	MOS	INT	MOS	INT	MOS	INT
PCM	3.81	0.45	3.89	0.13	4.16	0.13
PCM 2A	3.99	0.13	4.10	0.13	3.90	0.14
PCM 4A	3.35	0.12	4.02	0.14	3.70	0.14
PCM 8A	3.39	0.14	3.48	0.14	3.46	0.12
PCM 10^{-4}	3.31	0.15	3.55	0.14	3.15	0.16
PCM 10^{-3}	1.90	0.15	1.78	0.13	2.10	0.17
PCM + 10 dB	3.94	0.15	4.02	0.12	4.14	0.11
PCM – 15 dB	3.49	0.16	3.60	0.14	3.41	0.16
ADPCM	3.60	0.15	3.41	0.13	3.65	0.12
ADPCM 2A	3.72	0.13	3.30	0.12	3.38	0.13
ADPCM 4A	3.14	0.13	2.85	0.13	2.63	0.13
ADPCM 8A	2.51	0.14	2.09	0.14	2.23	0.15
ADPCM 2T	3.77	0.12	3.33	0.13	3.42	0.13
ADPCM 4T	3.86	0.14	3.01	0.14	3.80	0.13
ADPCM 10^{-4}	3.54	0.11	3.28	0.12	2.81	0.15
ADPCM 10^{-3}	2.88	0.16	2.55	0.15	1.93	0.13
ADPCM + 10 dB	3.80	0.14	3.55	0.14	3.61	0.13
ADPCM – 15 dB	3.20	0.15	3.02	0.15	2.92	0.14
ADPCM, C 2A	2.44	0.16	2.62	0.16	2.23	0.14
ADPCM, C 4A	2.13	0.15	2.14	0.13	1.90	0.13
ADPCM, C 8A	1.98	0.14	1.84	0.13	1.59	0.12

S/N 40	3.52	0.15
S/N 35	3.18	0.17
S/N 25	2.04	0.15
S/N 15	1.23	0.09
Q 10	1.41	0.10
Q 15	2.34	0.11
Q 20	3.04	0.10
Q 25	3.61	0.09
Q 30	3.96	0.09

Note 1 – Votes combined across four speakers and two sentences.

Note 2 – Number of votes = 128 except for Q where N = 256.

TABLE B-2

Degradation mean opinion scores (DMOS) and 95% confidence intervals (INT) for DCR test

Test conditions	X		Y		Z	
	DMOS	INT	DMOS	INT	DMOS	INT
PCM	4.35	0.10			4.41	0.11
PCM 8A	3.48	0.16			3.33	0.15
PCM 10^{-3}	2.21	0.11			2.25	0.14
ADPCM	4.33	0.11	4.22	0.11	4.05	0.12
ADPCM 8A	2.63	0.14	2.35	0.14	2.38	0.17
ADPCM 10^{-3}	3.14	0.16	2.83	0.14	1.85	0.14
ADPCM 4T	4.29	0.10	3.69	0.14	4.09	0.13

Q 15	1.99	0.15
Q 20	2.97	0.17
Q 25	3.89	0.18
Q 30	4.66	0.10
Q 35	4.81	0.09
Origin	4.98	0.03

Note 1 – Votes combined across four speakers and two sentences.

Note 2 – Number of votes = 128.

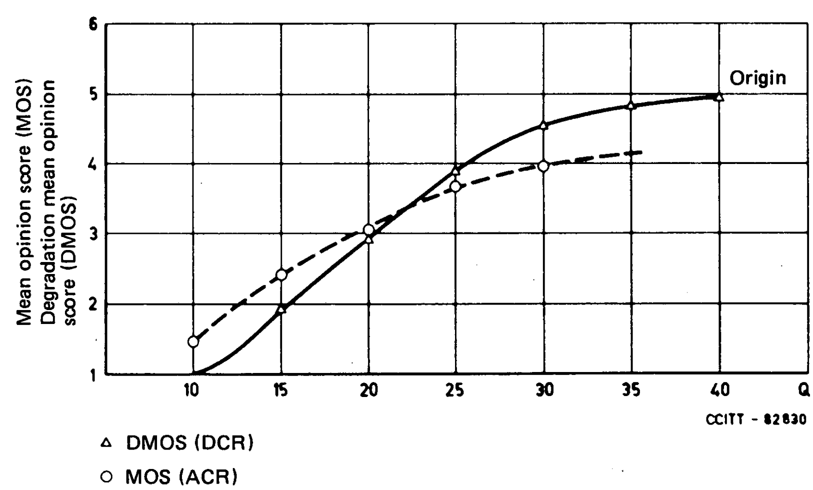


FIGURE B-1

MNRU reference conditions

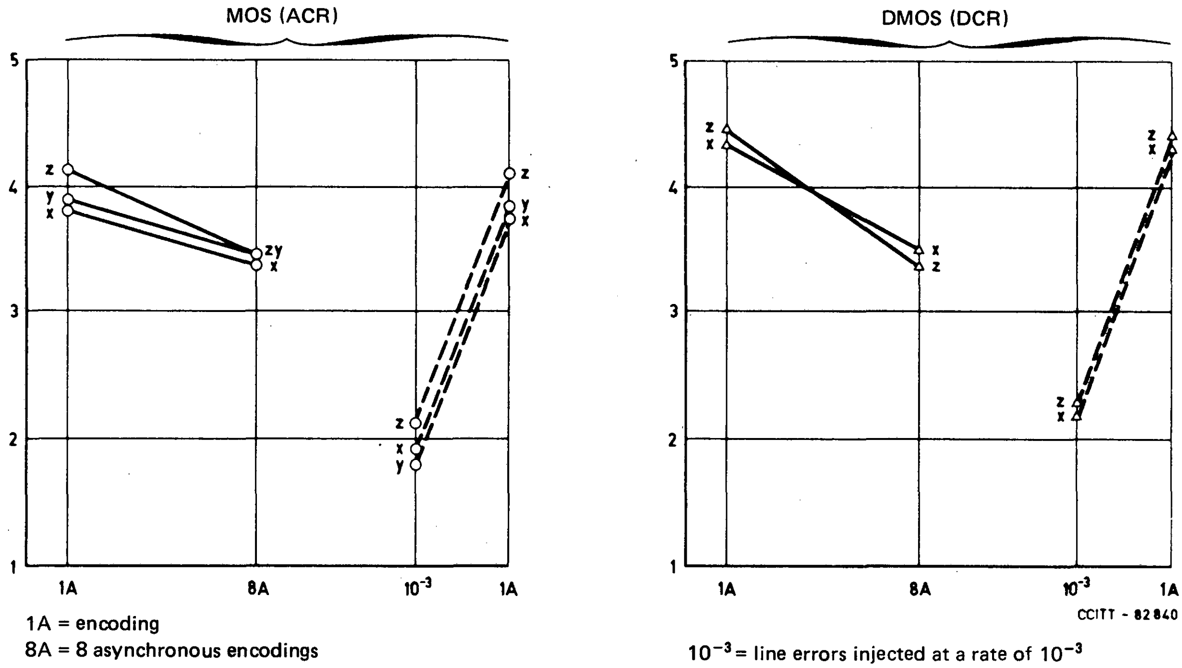


FIGURE B-2

PCM reference conditions

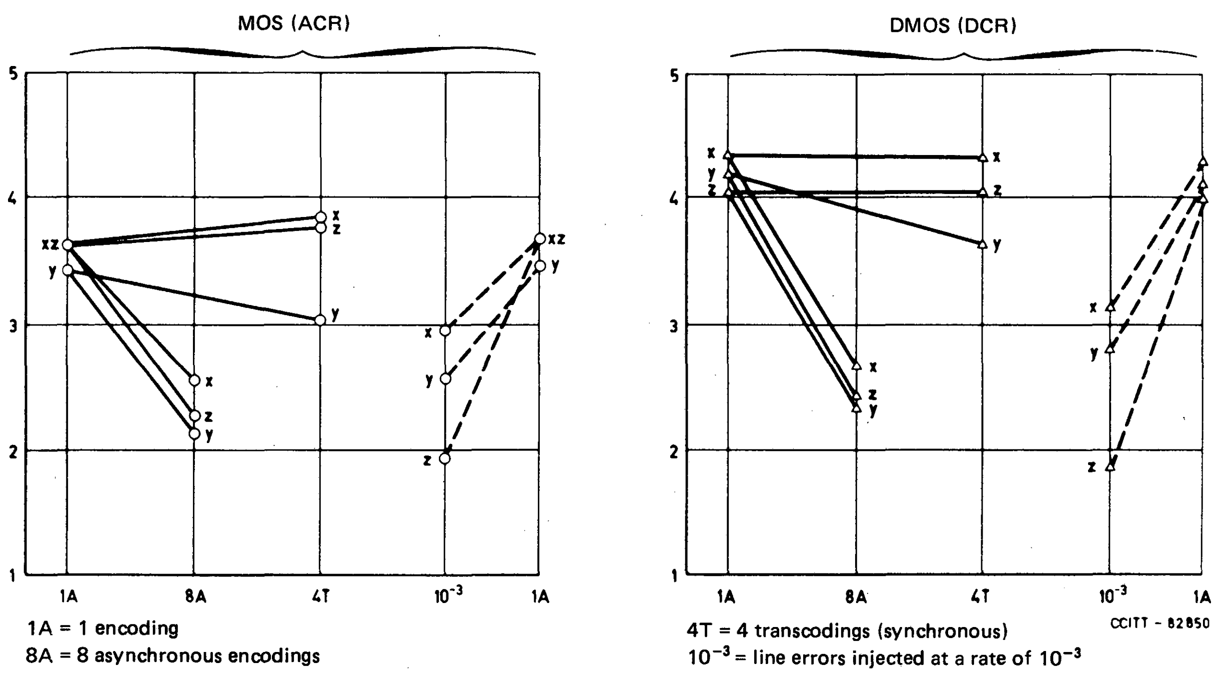


FIGURE B-3

ADPCM coders conditions

ANNEX C

(to Supplement No. 14)

Threshold method for direct comparison of digital encoders with a modulated noise reference unit (MNRU)

C.1 *Introduction*

By direct comparison of a digital system with an MNRU it is possible to assess the Q value which equals the performance of the system under test. The method described here leads to a threshold of equality defined as the 50% preference level between the MNRU and the digital system.

The threshold method is expected to give stable and precise results even for high quality digital processes. For wideband digital encoders the use of a wideband MNRU as described in Annex A of Recommendation P.81 is recommended.

C.2 *Testing procedure*

A listening-only test procedure is used. A signal pair consisting of a reference signal and a test signal is presented to listeners, who are then asked to indicate which of the signals in the pair they judge to have the highest quality (preference rating). Subjective equivalent SNR (Q) is defined as the reference SNR corresponding to the intersection point of the regression curve of the preference scores at the 50% preference level. An example of Q obtained with hypothetical preference scores is shown in Figure C-1.

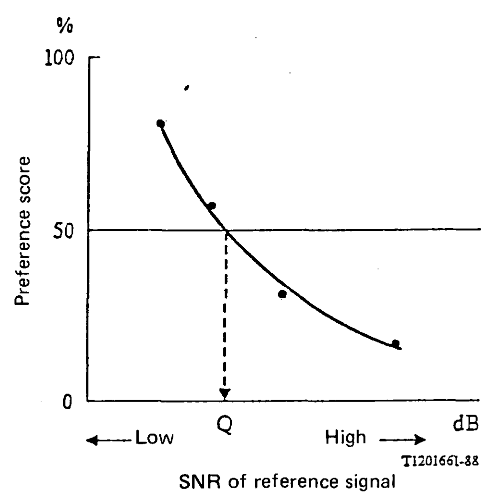


FIGURE C-1

Example of Q with hypothetical preference score

C.3 *Presentation of signals*

Reference signal A and test signal B are arranged in an equal number of A-B pairs and B-A pairs, and presented in random order. Several distortion levels spaced, for example, at 2 dB intervals, are introduced to the reference signal so that the range of preference scores extends from 20% to 80%, where the 50% preference lies in the middle of the distortion range. A timing diagram of the presentation is shown in Figure C-2.

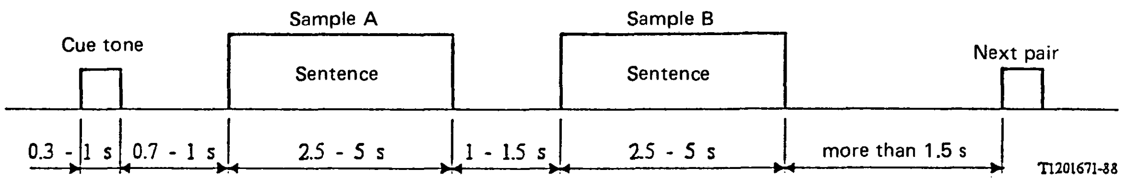


FIGURE C-2

Timing diagram of the presentation

The subject is required to make a judgement and respond by saying "A is better" or "B is better" (forced choice). The response "A equals B", or "No difference" is forbidden. The duration of the presentation should be limited to about six minutes in order not to tire the listeners. More listening samples may be presented after a suitable rest period. At least two, preferably four or five replications (repetitions of identical presentations) are recommended.

Note – If the MNRU is available in hardware and the SNR can be easily changed between presentations, a simplified procedure can be used. In this case the balancing to equally perceived quality is done by the subject. The adjustment is made during the pause between the pairs. The reference is always presented first. Presentation continues until the subject reports that the equality threshold has been reached.

C.4 *Speech sources*

It is necessary to use short sentences spoken by at least two males and two females, preferably four or six of each; different sentences are required for each speaker. The duration should be 2.5-5 seconds for speech and less than 10-15 seconds for music signals. Clicks at the beginning and end of the samples must be avoided. A linear microphone of sufficient bandwidth should be used to record the source signals in a sound-absorbent room having an ambient noise of less than 20 dBA and a reverberation time of less than 0.3 seconds in the band 125-8000 Hz. If digital recording equipment is used, the quantizing noise level should be less than the noise level in 14-bit linear PCM.

C.5 *Listening environment*

A high-fidelity sound reproduction system should be used for the listening test. When listening is carried out with loudspeakers, the reproduction equipment should be of studio-quality and the listening room should conform to CCIR Report 797. If headphones are used, diotic (binaural) listening is preferable. The bandwidth should be at least as wide as that of the digital system under test.

C.6 *Listeners*

Although it is preferred that listeners should be selected according to the description in the ACR method (see § A.4.1), this is not a strict condition in the pair comparison test. If the purpose of the listening test is to obtain the opinions of untrained listeners, untrained subjects are necessary. However, if this is not the purpose of the test, then trained listeners can be used and the reliability of the listening test can be extended by increasing the number of replications for each listener. The minimum number of listeners is six, but preferably twelve or more. Several subjects may listen simultaneously but it must be ensured that their responses are obtained independently.

C.7 *Reliability*

Since variations in preference score in subjective tests are assumed to conform to a t-distribution, the score variation width r which yields 95% reliability at score u ($0 \leqslant u \leqslant 1$) over the number of trials (i.e. the number of repetitions for each presentation pair multiplied by the number of subjects number of source signals) is presented in equation (C-1).

$$r = \pm\, t(n - 1, 0.05) \cdot \sqrt{u(1 - u)/(n - 1)} \qquad \text{(C-1)}$$

If n equals 96 and u equals 0.5 (preference score is 50%), r equals ± 10%.

ANNEX D

(to Supplement No. 14)

Principles of a method used by one organization in analyzing digital codec performance (Bell Communications Research, Inc.)

Hypothetical mean opinion score (MOS) results obtained from a subjective test conducted according to the methodology of Annex A are shown in Figures 1 to 4 of this Supplement. (Straight lines are used simply to connect data points.) Generally such results will display a saturation effect at and near the very good conditions (high MOS) and the very bad conditions (low MOS). (For high MOS, the saturation is caused by the 5-point scale and possibly by the idle circuit noise of the subjective test system without added impairments, e.g. idle circuit noise, Q, codec quantization distortion. For low MOS, the saturation is caused by the 5-point scale.)

An analytic method of data analysis used by Bell Communications Research, Inc. provides a value called "fit mean" for each test condition [8]. The fit means are then used in analysis of the data. (Fit means and MOSs are nearly equal over the mid-range of MOS values; however, fit means are not numerically constrained for extremely good and extremely bad conditions as are MOSs.) Plots of test results in terms of fit means will be similar to those of Figures 1 to 4, but will numerically exhibit greater spread.

The objective of the analysis is to determine a function:

$$Q_s = f(L) \tag{D-1}$$

where

Q_s = Q value for the codec quantization distortion,

L = Line bit rate (e.g. in kbit/s).

(A simple linear relation may be possible in some cases while other cases may require a more complex function.)

The codec Q can then be estimated from Equation (D-1).

Determination of Equation (D-1) needs to take into account in an appropriate manner the saturation effects discussed earlier. For example, the codec design line bit rate may correspond to the middle data point of Figure 3 for which there appears to be a modest saturation effect.

Similarly, the equivalence function (SNR *vs* Q) internal to the test may need to be considered. [This function is determined from appropriate functions fitted to the curves (in terms of fit means) similar to the curves of Figures 1 and 2.]

An important consideration in the analysis method is obtaining predicted performance values (fit means, Q) approximating as closely as possible the actual performance values (fit means similar in form to the curves of Figures 3 and 4 or Q values obtained by converting the fit mean values to Q values using an appropriate function fitted to fit mean data similar in form to the data of Figure 2). (For present purposes, it is assumed that for asynchronously tandemed codecs the combining law is 15 $\log_{10} n$, where n is the number of tandemed identical codecs. It may also be desirable to include the determination of the combining law in the analysis.)

The Q values obtained for a digital process according to the procedure described above can be used in various ways to assess the effect of quantization distortion on telephone connection performance. The subjective opinion model of Supplement No. 3 is in terms of corrected reference equivalent and idle circuit noise level. This model requires that the Q for an overall connection be converted to an equivalent SNR which can then be converted to an idle circuit noise level based on knowledge of the speech levels for connections of interest.

Equivalence functions used for this purpose have been found to vary (see [9], [10], [11], [12] and [13]). It is not clear if there exists a unique equivalence function which can be agreed on, and what that equivalence function should be. (Perhaps a basic equivalence function should be based on conversational test results.)

The Q values obtained for digital processes can also be used as a basis for specifying codec network application rules in terms of the number of asychronously tandemed 8-bit, μ-255 codecs. For the model of Supplement No. 3, the relation between the number of such codecs and Q (based on a $15 \log_{10} n$ law) is as follows:

Number [2]	1	2	3	4	5	6	8	10	12	14
Q(dB)	37	32.5	30	28	26.5	25.5	23.5	22	21	20

According to the model of Supplement No. 3, a 7-bit, μ-255 PCM codec would correspond to about 2.5 asynchronously tandemed 8-bit, μ-255 codecs. (Note that the value of Q = 37 dB is 3-4 dB greater than the minimum S/D values of Recommendation G.712 [15]; it is assumed that average 8-bit systems perform at the higher value.)

References

[1] CCITT Recommendation *32 kbit/s adaptive differential pulse code modulation (ADPCM),* Vol. III, Rec. G.721, ITU.

[2] GOODMAN (D. J.), NASH (R. D.): Subjective quality of the same speech transmission conditions in seven different countries, *Proc. ICASSP 82* (International Conference on Acoustics, Speech and Signal Processing), Vol. 2, Paris, May 1982.

[3] DAUMER (W. R.), CAVANAUGH (J. R.): A subjective comparison of selected digital codecs for speech, *Bell System Technical Journal,* Vol. 57, No. 9, November 1978.

[4] RICHARDS (D. L.), BARNES (G. J.): Pay-off between quantizing distortion and injected circuit noise, *Proc. ICASSP 82* (International Conference on Acoustics, Speech and Signal Processing), Vol. 2, Paris, May 1982.

[5] COMBESCURE (P.) *et al.* Quality evaluation of speech coded at 32 kbit/s by means of degradation category ratings, *Proc. ICASSP 82* (International Conference on Acoustics, Speech and Signal Processing), Vol. 2, Paris, May 1982.

[6] CCIR Doc. 11/17 *Subjective assessment of the quality of television pictures* (EBU), Study Period 1978-1982.

[7] TUCKEY: The problem of multiple comparisons, *Ditton,* Princeton University, Ed. 1953.

[8] CAVANAUGH (J. R.), HATCH (R. W.), SULLIVAN (J. L.) Models for the subjective effects of loss, noise and echo on telephone connections, *Bell System Technical Journal,* Vol. 55, November 1976.

[9] CCITT Contribution COM XII-No. 24 (Study on determination of subjectively equivalent noise (NTT)), Study Period 1981-1984.

[10] CCITT Contribution COM XII-No. 61 (Subjectively equivalent noise for linear and carbon microphone originated speech signals – Bell Northern Research, Canada), Study Period 1981-1984.

[11] CCITT Contribution COM XII-No. 124 (Application of information index to quantizing noise in PCM – France), Study Period 1981-1984.

[12] CCITT Contribution COM XII-No. 130 (Subjective equivalence functions for consideration in a quantization distortion opinion model – Bell Northern Research, Canada), Study Period 1981-1984.

[13] CCITT Contribution COM XII-No. 162 (Transmission performance of digital systems – COMSAT), Study Period 1981-1984.

[14] CCITT Recommendation *Transmission impairments,* Yellow Book, Vol. III.1, Rec. G.113, ITU, Geneva, 1981.

[15] CCITT Recommendation *Performance characteristics of PCM channels audio frequencies,* Yellow Book, Vol. III.3, Rec. G.712, ITU, Geneva, 1981.

[16] CCITT Contribution COM XII-No. 222 (Evaluation of speech transmission quality by ACR and DCR methods – USSR Telecommunications Administration), Study Period 1985-1988.

[17] CCITT Contribution COM XII-No. 223 (Results of determination of typical dependence of average subscriber's evaluation from Q values – USSR Telecommunications Administration) Study Period 1985-1988.

[18] CCITT Contribution COM XII-R5, page 49 and pages 76-78, June 1985.

[2] This should not be interpreted as the number of qdus [14] to be used for international planning; the relation between the number of codecs and Q applies for the model of Supplement No. 3 which has been used in planning studies in the United States.

WIDEBAND (7 kHz) MODULATED NOISE REFERENCE UNIT (MNRU) WITH NOISE SHAPING

(Melbourne, 1988)

(Quoted in Recommendation P.81)

(Contribution of NTT Japan)

1 Introduction

The configuration of a wideband MNRU takes into account the following three points with respect to its use as a common reference signal:

a) The procedure for generating the reference signal should be simple and clear;

b) The speech quality characteristics of the reference signal should be similar to those of the test signals;

c) It should be possible to control grades of degradation arbitrarily.

Compared to the narrow-band MNRU, the wideband MNRU has an enlarged bandwidth (70-7000 Hz) and a fixed noise spectrum shaping filter which lessens the high-frequency range noise, thus making the noise spectrum in the reference signal resemble that in wideband encoders.

2 Arrangement of the wideband MNRU

The basic arrangement of the wideband MNRU is shown in Figure 1. Wideband Gaussian noise instantaneously multiplied by source speech is fed to a spectrum shaping filter. The source speech and the shaped-spectrum noise are band-limited and attenuated/amplified to obtain the desired SNR, and then both are added to produce the distorted signal.

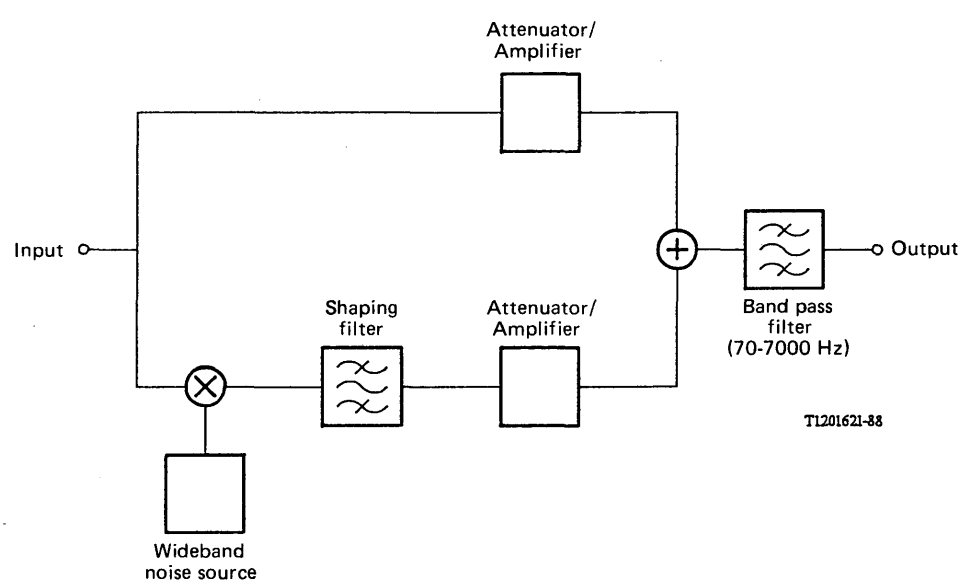


FIGURE 1

Arrangement of the wideband MNRU with noise shaping

3 Spectrum shaping filter

The noise spectrum is shaped with a 1st order auto-regressive filter, whose diagram is shown in Figure 2. A computer simulation using a sampling frequency of 16 kHz and a bandwidth of 7 kHz yields a filter coefficient of 0,8 which approximates the long-term speech spectrum envelope of wideband source signals.

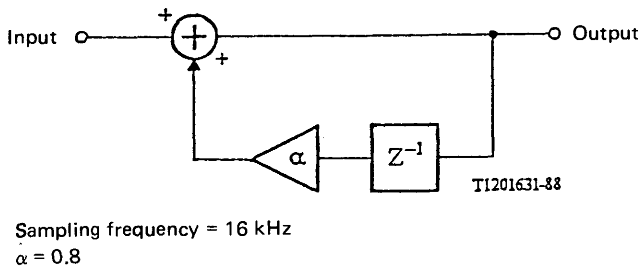


FIGURE 2

Noise shaping filter

4 Band-pass filter

The bandwidth of the wideband MNRU should correspond to that of wideband speech encoders. The provisional frequency response requirements for the 7 kHz band-pass filter are shown in Figure 3 based upon the present output filter in Recommendation P.81.

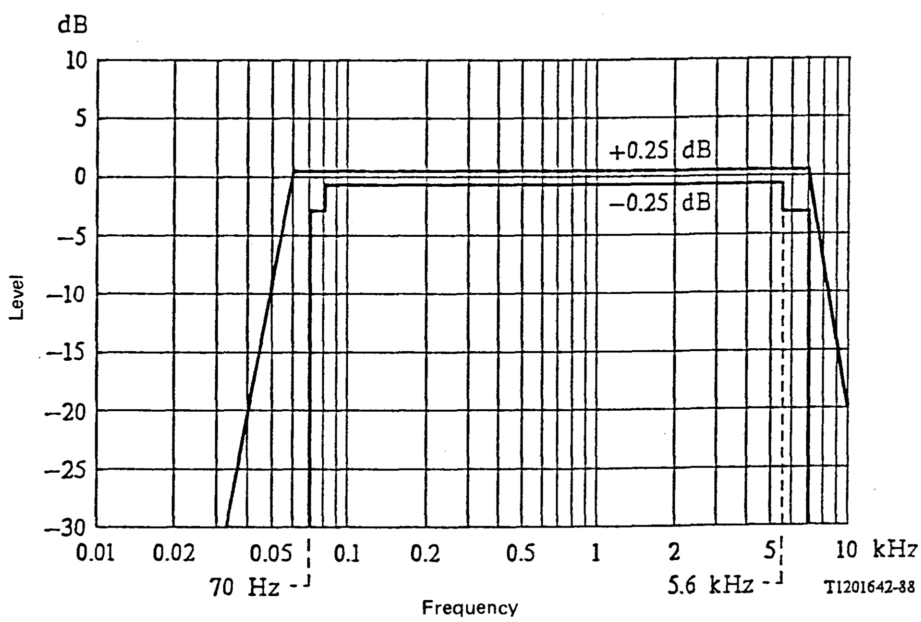


FIGURE 3

Requirements for the output filter used in the wideband MNRU

5 Signal-to-noise ratio

In a computer simulation, SNR can be calculated using mean power with a time constant of 8-20 milliseconds after band-pass filtering of both the source speech and noise. When the SNR is set using a sinusoidal source signal for the MNRU equipment, measurement of noise power should actually be carried out using an r.m.s. volt meter, rather than by simply calibrating the loss or gain of the noise channel based on the sinusoidal signal.

6 Other specifications

Sampling frequency: It is recommended that the sampling frequency be more than twice the upper pass-band frequency; with respect to the test of Recommendation G.722 encoders, a sampling frequency of 16 kHz is recommended.

Noise source: For MNRU equipment, sufficiently wideband Gaussian noise should be used. If the noise is computer-generated, a full flat spectrum over one half the sampling frequency band is necessary. Amplitudes greater than three r.m.s. value should be clipped.

Supplement No. 16 [1)]

GUIDELINES FOR PLACEMENT OF MICROPHONES AND LOUDSPEAKERS IN TELEPHONE CONFERENCE ROOMS [1] AND FOR GROUP AUDIO TERMINALS (GATs)

(Malaga-Torremolinos, 1984; amended Melbourne, 1988)

(Quoted in Recommendations G.172 and P.30)

1 General

The following guidelines provide basic rules for assessing the acoustics of telephone conference rooms and for installing group audio terminals consistent with maximum speech intelligibility and easy talker recognition.

2 Conference room acoustics – General requirements

The design and installation of a telephone conferencing system or group audio terminals which meet reasonable cost and performance specifications involve numerous judgements and trade-offs. These guidelines will enable the planner and installation engineer to assess the acoustics of a room, to make the necessary choices and decisions, to install the appropriate equipment properly and thereby provide satisfactory service.

The audio portion of a group audio terminal consists of terminal equipment with microphones and loudspeakers installed in conference rooms and interconnected by an audio transmission facility. This transmission facility may be either public switched telephone connections or private line facilities.

In both public and private systems, transmission is frequently interconnected via a multipoint conference bridge so that each room can communicate simultaneously with any of the other locations. When this is done, it is most important that the bridge be located at the electrical loss center of the network in order to minimize level contrast between the speech originating in the different rooms.

1) Formerly supplement No. 25 to Fascicle III.1 *(Red Book).*

Unlike telephony between handsets, the acoustic properties of the conference room and the placement of microphones in the room critically determine the level, the speech signal-to-ambient-noise ratio and the reverberant quality (rain-barrel effect) of the transmitted speech. Particularly in multipoint conferences, these three factors are easily judged and critically commented upon by users.

In general, the larger, noisier and more reverberant a room is, the less suitable it will be for group communication. The presence of noise and/or reverberation in the transmitted speech results in a system whose performance is unsatisfactory. In extreme cases, experience has demonstrated that excess noise in one room, e.g. from overflying aircraft, can temporarily block transmission between all rooms of a multipoint system. Excess reverberation results in such hollowness to the received speech that talkers become difficult to recognize and understand, causing users to fatigue easily and refuse to use the system.

In principle, any room is suitable for group communication if these guidelines are followed. However, the guidelines will dictate that in a noisy or reverberant room, talkers must speak so close to microphones that they might as well use handsets. The user requesting the installation must then choose one or more of the following options:

1) select another conference room;
2) acoustically treat the room; or
3) accept the close microphone/talker distances dictated by the guidelines.

Several very important criteria must be fulfilled simultaneously to assure satisfactory audio performance of a telephone conference system. The balance of this section describes the determination of these criteria. Briefly, these criteria are:

1) A room suitable for a normal face-to-face conference must be selected.
2) A noise dependent microphone/talker distance must be determined.
3) A reverberation dependent microphone/talker distance must be determined.
4) The microphones and loudspeakers must be positioned in accordance with both these distances.

3 Ambient noise level considerations

The ambient noise level requirements for conference rooms of increasing size and number of conferences are given in Table 1. As the room size and number of conferences increase, the participants will sit further apart. Consequently, for comfortable talking and listening, the ambient noise level in the room must decrease as the group size increases.

TABLE 1

Ambient noise level limits for conference rooms

Room description	Maximum sound level meter reading	Acoustic environment
Conference room for 50 people	35	Very quiet, suitable for large conferences at tables 6-9 m in length
Conference room for 20 people	40	Quiet, satisfactory for conferences at tables 4.5 m in length
Conference room for 10 people	45	Satisfactory for conferences at tables 1.5-2.5 m in length
Conference room for 6 people	50	Satisfactory for conferences at tables 1.0-1.5 m in length

Noise measurements as stipulated in Table 1 should be performed at the conference table with the room in normal operation but unoccupied. These noise measurements should be performed at least 0.6 m away from any surface.

Noise measurements in dBA can be made with a sound level meter employing A-weighting, a reference pressure level of 20 μ Pascal and otherwise conforming to Recommendation P.54. A-weighting is used in these guidelines since it approximates the annoyance level of noise to the human ear.

The maximum microphone/talker distance is limited by ambient noise. Figure 1 shows the maximum distance between a talker and a microphone which ensures a marginally acceptable signal-to-noise ratio of 20 dB in the transmitted speech. No attempt should be made to ignore or increase this distance beyond that determined in Figure 1. As an example, with an ambient noise level of 50 dBA, Figure 1 shows that the *maximum* distance (D_{max}) from talker to microphone for *marginal* acceptability is 0.5 m. Figure 1 applies to omnidirectional microphones. When directional microphones, e.g., cardioid or bidirectional are used, the D_{max} value determined in Figure 1 can be increased by 50 percent.

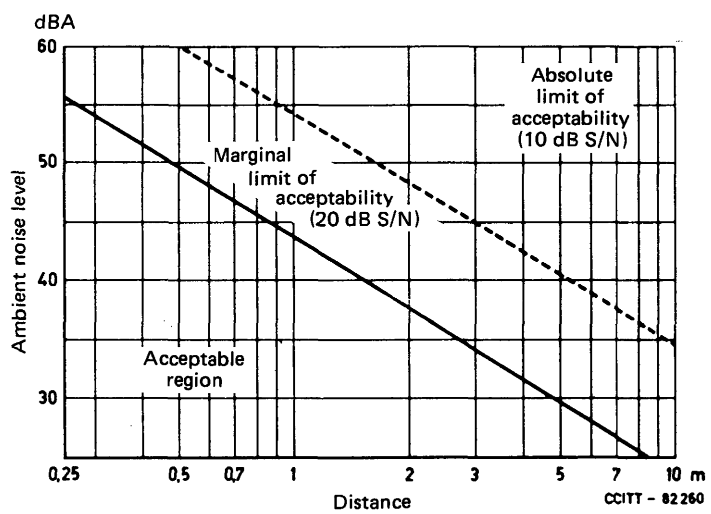


FIGURE 1

Maximum distance from talker to microphone

If more than one microphone is used to cover more than two or three talkers, and all microphones are active at the same time, then the amount of room noise picked up by the microphones and transmitted on the circuit will increase. How much it will increase is not completely predictable but a useful approximation is that the apparent noise level will rise 3 dB each time the number of microphones is doubled. This apparent rise in the effective noise level can be taken into account by adding it to the measured noise level before using Figure 1 to determine D_{max}.

4 Reverberation considerations

Most rooms for telephone conferencing have acoustical characteristics which cannot be altered, thus the quality of sound transmitted from the room can only be controlled by microphone placement. When the microphone is close to the talker, the greatest percentage of sound picked up comes directly from the talker, reverberation in the room would exert relatively little influence. As the distance between the microphone and the talker increases, the direct sound level reaching the microphone decreases 6 dB for each doubling of the distance, whereas the average level of the reverberant sound remains more nearly constant.

The critical distance (D_c) of a room is a useful concept to describe a room. It is the distance from a sound source (talker, loudspeaker) at which the direct sound energy from the source equals the reverberant energy reflected off all room surfaces (walls, ceiling, furnishings, floor). Critical distances in conference rooms are typically in the range of 0.2 to 1.5 meters.

The critical distance can be expressed as:

$$D_c = 0.056 \sqrt{\frac{V}{T_R}} \text{ meters} \qquad \text{(see ISO 35u)}$$

where

V is the volume of the room in cubic meters,

T_R is the reverberation time of room in seconds.

As the ratio of direct-to-reverberant sound energy decreases with increasing microphone/talker separation, reproduced speech becomes less intelligible, of poorer quality, difficult to recognize and fatiguing to listen to. It acquires a hollowness which sounds as if the person were speaking from the bottom of a rain-barrel. For good performance, microphones should be placed at no more than half the critical distance (0.5 D_c) from talkers. This usually requires installing multiple microphones on the conference table or lavaliere microphones[2)] on conferees, and definitely rules out placing microphones in the ceiling. Many installations for group communication have failed because microphones were installed in ceilings without regard to the above acoustic requirements.

When directional (cardioid or bidirectional) microphones are used, the distance between microphones and talkers may be increased by 50 percent, to three-quarters of the critical distance (0.75 D_c). For best results, talkers must sit in front of cardioid (heart shape) microphones; they may sit on either side of a vertically mounted bidirectional microphone with a cosine (figure-eight shape) sensitivity pattern. Table 2 gives typical microphone/talker separation distances for small (60-300 m^2 of wall, ceiling and floor surface area) and large (300-1000 m^2) rectangular conference rooms, together with the estimated critical distance (D_c). Areas in square meters are used in these guidelines, since they are much more relevant to conference room acoustics than are the often quoted room volumes.

TABLE 2

Typical microphone/talker separation (meters)

Conference room	Omnidirectional microphone	Directional microphone	Critical distance
Small room (60-300 m^2)			
moderate room treatment [a)]	0.3	0.5	0.6
Large room (300-1000 m^2)			
some room treatment [a)]	0.6	0.9	1.2
considerable room treatment [a)]	0.9	1.4	1.8

[a)] In this context, a room with moderate treatment might have an accoustic ceiling and a carpet on the floor; one with some treatment might have either an accoustic ceiling or a carpet; while a room with considerable treatment might have heavy, lined drapes covering half the wall area in addition to a high-quality suspended acoustic ceiling and a thick carpet with underfelt.

[2)] Microphones with an attached, adjustable strap which can be hung around the neck of the user.

5 Microphone type and placement

As stated earlier, when omnidirectional microphones are used the microphone/talker distance must be less than the maximum distance (D_{max}) determined from Figure 1 to ensure adequate signal-to-noise ratio. When directional microphones are used, the microphone/talker distance can be increased but must be less than 1.5 D_{max}.

Also stated earlier, when using omnidirectional microphones, the microphone/talker separation must be less than half the critical distance to ensure highly-intelligible, easily-recognizable, nonreverberant speech. When directional microphones are used, the microphone/talker distance can be increased but must be less than 0.75 D_c.

Microphones must be placed to satisfy *both* the above rules; in other words the microphone/talker distance must not exceed the smaller distance.

So that all talkers can satisfy the above microphone/talker criteria, more than one microphone is usually required. Typically one microphone for every 3 talkers is necessary. For each doubling of the number of microphones, the effective noise level in the room will increase by 3 dB. Thus, in the example of § 3 if four microphones were used, the reading of 50 dBA would be raised to an effective value of 56 dBA. The noise determined, D_{max} from Figure 1 would thus be reduced to 25 cm. Clearly, lavaliere microphones would provide a practical solution to keeping talkers within 25 centimeters of a microphone.

6 Loudspeaker placement

The requirements for placing loudspeakers in a conference room are much less critical than those for microphones. It is generally considered good practice to limit the distance from any listener in the room to the nearest loudspeaker to not more than twice the critical distance.

Loudspeakers should be distributed in the ceiling, on the walls, or on the conference table to ensure a minimum sound pressure level of 65 dBA at listener positions. If there is significant noise, the sound pressure level should be at least 20 dB above the ambient noise level. More "presence" and less "voice-on-high" effect is achieved when the loudspeakers are placed on or in the edge of the conference table.

Ceiling mounted loudspeakers are usually simpler to install and less conspicuous. Generally, loudspeakers installed in a visible grid, suspended, acoustic panel ceiling should be placed approximately 0.6 meters outside the edge of the conference table. Best results are obtained when the loudspeakers are *not* installed symetrically but somewhat randomly. This prevents exciting pronounced room modes of vibration.

Reference

[1] *Teleconference center construction guidelines*, Bell System Technical Reference, PUB 42903, May 1980, American Telephone and Telegraph Co.

Supplement No. 17

DIRECT LOUDNESS BALANCE AGAINST THE INTERMEDIATE REFERENCE SYSTEM (IRS) FOR THE SUBJECTIVE DETERMINATION OF LOUDNESS RATINGS

(Melbourne, 1988)

(Quoted in Recommendations P.78)

(Contribution from China)

1 Introduction

In the subjective determination of loudness ratings according to Recommendation P.78, the wideband fundamental reference system NOSFER should be always used in addition to the Intermediate Reference System (IRS). The main reason for using the indirect method for the subjective determination of loudness ratings is the difficulty to hold two handsets, one of the IRS and the other of the unknown system, in one hand during balance.

Since 1982, the CCITT Laboratory and some other laboratories have tried to use the direct loudness balance method for the subjective determination of loudness ratings using a cut-out handset. Results show that not only can the test be simplified, but also the discrepancies of the test results can be reduced considerably. Typically the standard deviation of the test results is only half of that using the Recommendation P.78 technique. Furthermore, the introduction of NOSFER in the subjective determination of loudness ratings is no longer necessary.

This Supplement describes the essential arrangement used in the direct loudness balance method.

2 Method

2.1 *Handset*

The IRS sending handset with its microphone is mounted in a loudness rating guard-ring position (LRGP) support. However, the handle along with the microphone holder of the IRS receiving handset may be cut away, if necessary, to facilitate holding both an unknown handset and the IRS cut-out receiver piece in one hand during the subjective balance for the RLR or OLR.

2.2 *Speech volume*

Experiments show that the average reading of a VU meter connected to the output of the IRS sending system is about -1.7 dB while an operator is speaking into the microphone of the IRS sending handset at the LRGP using the "standard volume" (see Recommendation P.72, Red Book). This value will be different if a different volume meter is used. Experiment results show that it is not necessary to establish the individual relationship between the "standard volume" and the reading of a meter connected to the output of the IRS sending system for each of the operators.

Because the bandwidth of the IRS sending system is limited, the fluctuation of the needle of the meter is larger than in the case of a wideband system while the talker is active. However, it is not difficult for the operator to control his volume within 1 or 2 dB using his own rule of reading.

2.3 *Listening level*

The loss inserted into the overall IRS connection is fixed at 18 dB, because this value is close to the "X2" value (refer to Recommendation P.78) determined by the recent subjective test team of the CCITT Laboratory as well as those of other laboratories.

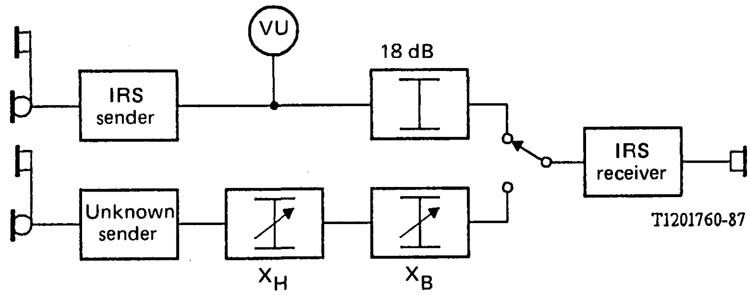
2.4 *Test arrangements*

The test arrangements for the determination of SLR, RLR, OLR and JLR are shown in Figure 1 to Figure 4.

2.5 *Balance method*

The "margin" method is used. The details are similar to the subjective determination of R25 equivalent, see Recommendation P.72 (Red Book).

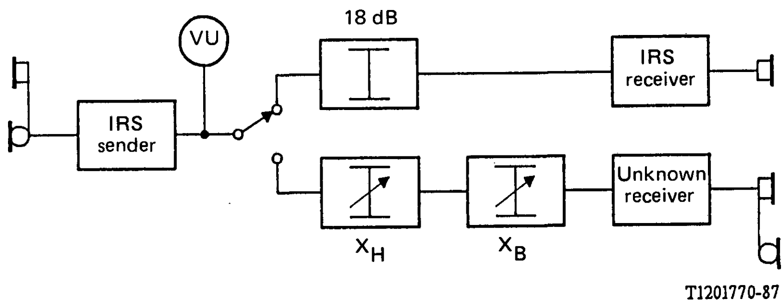
In the determination of RLR and OLR, the operator tends automatically to apply more force to the cut handset fitted to his ear because he holds the cut handset by his fingers directly, while at the same time holding the handle part of the "unknown" handset. This is why the test results of RLR and OLR found in some laboratories are about 1 to 2 dB larger (quieter) than those using the Recommendation P.78 method. This effect can be eliminated if the operator is told that his ear must feel the same force whether the earcap of the handset of an "unknown" system or the earcap of the cut handset of the IRS is applied to his ear.



$SLR = 18 - (X_B + X_H)$

FIGURE 1

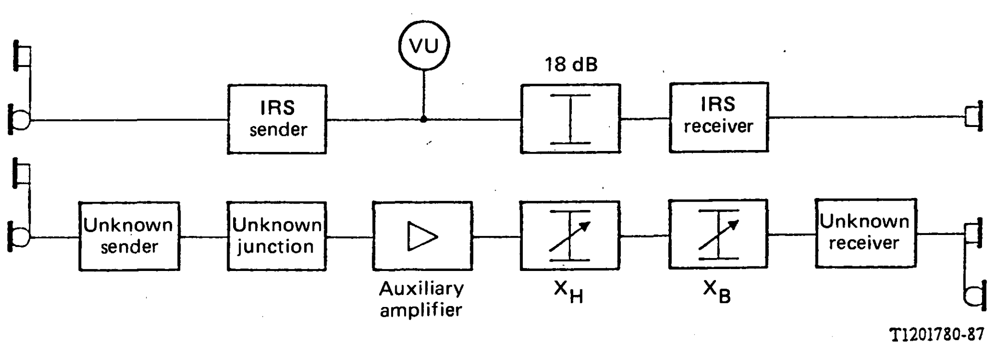
Determination of SLR



$RLR = 18 - (X_B + X_H)$

FIGURE 2

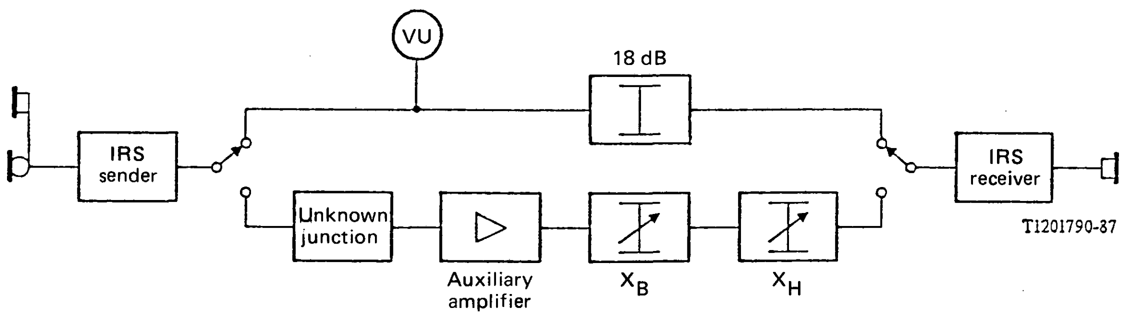
Determination of RLR



$OLR = 18 - (X_B + X_H)$ + Gain of the auxiliary amplifier

FIGURE 3

Determination of OLR



JLR = 18 – (X_B + X_H) + Gain of the auxiliary amplifier

FIGURE 4

Determination of JLR

Supplement No. 18

COMPARISON OF THE READINGS GIVEN ON SPEECH BY METERS CONFORMING TO RECOMMENDATIONS P.52 AND P.56

1 Introduction

This Supplement gives information on the internationally coordinated "round-robin" experiment which compared the readings of standard recordings given by VU [1], ARAEN [2] and peak programme [3] meters conforming to Recommendation P.52 and the speech voltmeter conforming to Recommendation P.56.

2 Composition of the tapes

The recorded material was made in three languages: English, Polish and Singhalese. Male and female talkers were also required in order to give a wide range of frequencies and timbre of voice.

Each talker recorded a list of 5 short sentences with approximately a 5-second gap between each sentence and the next. This was followed by a short passage of prose of aboute 1 minute's duration. Each pair of talkers then held a conversation in their mother tongue lasting several minutes. To stimulate the conversation, each talker was given a standard set of picture cards normally used in conversation experiments (Supplement No. 2).

The speech levels recorded were not altered because the tapes were intended to test the capability of the various meters and methods used in measuring. In practice this gave a range of about 15 dB.

3 Frequency responses

Two frequency responses were chosen for this experiment: the IRS sending end (Recommendation P.48) and a wideband flat response (±0.5 dB over the frequency range 100 to 6300 Hz).

4 Information about the tapes

The tapes were prepared on a 2-track (2-channel) tape recorder with IEC equalisation at a speed of 7.5 ips.

The format of each track was as follows:

At the beginning of each tape there was a 1 kHz tone that lasted for approximately 12 s (considered as the reference level). This was followed by a period of silence that lasted for approximately 5 s, which was then followed by the first speech condition to be measured. Each speech condition was followed by 5 s of silence and then 4 s of tone to indicate that the next speech condition began after a few more seconds of silence.

The contents of the tapes is shown in full in Table 1.

The output of the replay tape recorder was set in such a way that the level of the 1 kHz tone at the beginning of each track fell in the following range: 0 dBm to −10 dBm across 600 ohms (i.e. −2.2 dBV to −12.2 dBV).

5 Results

When used for measuring continuously spoken speech, the VU meter is specified to be read by taking the average of the peak deflections approximately every 10 seconds after excluding the two or three highest readings. For the ARAEN meter the readings are specified to be interpreted according to the CCITT rule of observing the reading which is exceeded on the average once every three seconds.

There is naturally an error due to the human element in interpreting any single reading.

The results from the "round-robin" experiment are shown in Table 2. No obvious differences were observed between the variables in the experiment, i.e. language, bandwidth, speech material and talkers, and therefore the results presented in Table 1 have been averaged over all variables, including observers. All readings from meters conforming to Recommendation P.52 are compared to the meter conforming to Recommendation P.56.

As can be seen from the results there is a wide variation. For the VU meter the range is some 6 dB. It must be borne in mind that this range is larger if the individual readings of the observers are taken into consideration, especially as some laboratories used more than one observer. In fact, the total range is slightly greater than 8 dB. These findings are consistent with those stated in [4].

In general, there appears to be consistency within a laboratory but inconsistency between laboratories. For example, for the USA there is an average difference between laboratories of nearly 6 dB.

The findings are similar for both the ARAEN meter and peak programme meter (PPM).

The results obtained from meters conforming to Recommendation P.56 showed, in general, a variation of less than 1 dB between all observers.

These results can be compared with older data from British Telecom and show that the "world" average for the VU meter agrees favourably, but for the ARAEN meter there is a difference of some 3 dB. However, the new results from British Telecom are consistent with this older data.

It is obvious that care is needed when comparing results between countries using meters conforming to Recommendation P.52 and the results from this experiment give guidelines to the differences to be expected.

TABLE 1

Contents of the tapes

Condition	Talker	Language	Speech material
Tape 1 – Track 1 (wideband)			
1 kHz tone			
1	male	english	short sentences
2	male	english	narrative
3	male	english	conversation
4	female	polish	short sentences
5	female	polish	narrative
6	female	polish	conversation
7	female	singhalese	short sentences
8	female	singhalese	narrative
9	female	singhalese	conversation
Tape 1 – Track 2 (wideband)			
1 kHz tone			
10	female	english	short sentences
11	female	english	narrative
12	female	english	conversation
13	male	polish	short sentences
14	male	polish	narrative
15	male	polish	conversation
16	male	singhalese	short sentences
17	male	singhalese	narrative
18	male	singhalese	conversation
Tape 2 – Track 1 (IRS sending)			
1 kHz tone			
19	male	english	short sentences
20	male	english	narrative
21	male	english	conversation
22	female	polish	short sentences
23	female	polish	narrative
24	female	polish	conversation
25	female	singhalese	short sentences
26	female	singhalese	narrative
27	female	singhalese	conversation
Tape 2 – Track 2 (IRS sending)			
1 kHz tone			
28	female	english	short sentences
29	female	english	narrative
30	female	english	conversation
31	male	polish	short sentences
32	male	polish	narrative
33	male	polish	conversation
34	male	singhalese	short sentences
35	male	singhalese	narrative
36	male	singhalese	conversation

TABLE 2

Comparison of readings made on meter conforming to P.52 and P.56

(Readings in dB relative to reading of meter conforming to Rêc. P.56)

Country	VU	ARAEN	PPM
USA			
AT&T	+4.5		
Bell Labs.	−1.2		
Sweden			
Telecom Admin	+0.7		
LME	−1.1	+0.3	
Australia			
Telecom Australia	+0.3		
Norway			
Telecom Admin	+0.4		
PR of China			
1sr Research Inst.	+2.9		
UK			
STL	−1.3	+1.4	
British Telecom		+5.0	+10.3
CCITT	0.0	+1.8	
France			
CNET	−2.1 [a] A-weighted +1.7 corrected		
Japan			
NTT		+3.0	
Average	+0.7 [b]	+2.3	+10.3

[a] French results were made with "A-weighting" and a correction was made to eliminate this effect.

[b] The "world" average used the corrected French result.

References

[1] CCITT *Volume Meter Standardised in the United States of America, Termec VU Meter,* Supplement No. 11, White Book, Volume V, 1969.

[2] CCITT *ARAEN Volume Meter or Speech Voltmeter,* Supplement No. 10, White Book, Volume V, 1969.

[3] CCITT *Modulation Meter Used by the British Broadcasting Corporation,* Supplement No. 12, White Book, Volume V, 1969.

[4] CCITT *Comparison of the readings given on conversational speech by different types of volume meter,* Supplement No. 14, White Book, Volume V, 1969.

[5] RICHARDS (D. L.): Telecommunication by speech, page 59, *Butterworths,* London, 1973.

INFORMATION ON SOME LOUDNESS LOSS RELATED RATINGS

(Melbourne, 1988)

(quoted in Recommendations P.79 and G.III)

Introduction

It is important to determine the electroacoustic performance of telephone sets in terms of a standard which is universally recognized. Recommendation P.79 gives the algorithm as agreed by the CCITT for calculation of loudness ratings (LRs) of telephone sets. In order to avoid confusion, this algorithm should not be changed during the 1989-1992 Study Period. However, it is also clear from several independent investigations that Recommendation P.79 represents only with limited accuracy the speech and hearing characteristics of "ordinary people". This Supplement, gives the reader of the P-Series Recommendations a possibility to study the back-ground to the problems and provides information on some other loudness rating systems which have been used.

In particular, §§ 1,7 and 8 give examples of algorithms found useful by some Administrations for their own national planning.

To avoid confusion when dealing with loudness ratings, the reader should also consider the information given in the *Preliminary Notes* to this Volume.

The results of (CCITT) LRs and R25Es calculated by the Chinese algorithm described in § 3 have been found to be in good agreement with the subjectively determined values obtained by the CCITT Laboratory in the past. This algorithm will be used by the CCITT Laboratory for the objective determination of R25Es for Administrations and other organizations.

1 **The IEEE algorithm for calculating "objective loudness ratings"** (Contribution from BNR, Canada)

Abstract

An algorithm for calculating loudness ratings is described. The algorithm is based on objective measurements and computations performed in such a manner that the numerical results obtained reflect the subjective attribute of loudness, but it employs certain simplifying assumptions, to combine simplicity and reasonably close agreement between objectively determined responses and subjective responses.

1.1 *Introduction*

The algorithm described below is based on a method [1] which has been in widespread use in North America for several years. The method has proved very adequate for use both in the planning of telephone networks and the characterization of individual components.

The method described may be used for determining the loudness rating of partial or complete connections. For complete connections, comprising overall or sidetone transmission paths, the procedure involves measurement of acoustic input and output pressures. For partial telephone connections comprising transmitting, receiving or electrical connection paths, the procedures involve measurement of acoustic pressure and electrical voltages. A particular advantage of this method for planning purposes, is that the sum of the loudness losses determined for individual parts of a connection closely approximates the loudness of the overall connection.

1.2 *Definitions*

1.2.1 **loudness rating**

The amount of frequency-independent gain that must be inserted into a system under test so that speech sounds from the system under test and a reference system are equal in loudness (see § 1.3.2).

1.2.2 reference system

A system that provides 0 dB acoustic gain between a mouth reference point at 25 mm in front of a talker's lips and an ear reference point at the entrance to the ear canal of a listener, when the listener is using an earphone. This system is assigned a loudness rating of 0 dB. The frequency characteristic of the system must be flat over the range 300-33000 Hz and show infinite attenuation outside of this range.

1.2.3 objective loudness rating (OLR)

The rating of a connection or its components when measured according to the methodology described within § 1.

1.2.4 overall objective loudness rating (OOLR)

$$OOLR = -20 \log_{10} \frac{S_E}{S_M} \tag{1-1}$$

where

S_M is the sound pressure at the mouth reference point (in pascals)

S_E is the pressure at the ear reference point (in pascals).

1.2.5 transmitting objective loudness rating (TOLR)

$$TOLR = -20 \log_{10} \frac{V_T}{S_M} \tag{1-2}$$

where

S_M is the sound pressure at the mouth reference point (in pascals)

V_T is the output voltage of the transmitting component (in millivolts).

1.2.6 receiving objective loudness rating (ROLR)

$$ROLR = -20 \log_{10} \frac{S_E}{\frac{1}{2} V_W} \tag{1-3}$$

where

V_W is the open-circuit voltage of the electric source (in millivolts)

S_E is the sound pressure at the ear reference point (in pasclas).

1.2.7 electrical objective loudness rating (EOLR)

For an electrical network,

$$EOLR = -20 \log_{10} \frac{V_T}{\frac{1}{2} V_W} \tag{1-4}$$

where

V_W is the open-circuit voltage of the electric source (in millivolts)

V_T is the output voltage of the network (in millivolts).

1.2.8 *Loudness equation*

Loudness voltages (in millivolts) and pressures (in pascals) are determined in accordance with Equation (1-5).

$$X = \left\{ \frac{\sum_{j=2}^{N} \left(\log_{10} \frac{f_j}{f_{j-1}}\right) \left[\frac{\left(10^{\frac{x_j}{20}}\right)^{1/2.2} + \left(10^{\frac{x_{j-1}}{20}}\right)^{1/2.2}}{2} \right]}{\log_{10} f_N / f_1} \right\}^{2.2} \tag{1-5}$$

where

x_j is the signal level of S_E, S_M, V_W or V_T (in dBPa or dBmV) at frequency f_j

X the value S_E, S_M, V_W or V_T in accordance with the value used for X_j

f_j are specific frequencies of the N frequencies selected for analysis.

Loudness voltages and pressures are expressed in decibel-like form using Equation (1-6).

$$S'_E,\ S'_M,\ V'_W \text{ or } V'_T = 20 \log_{10} X \tag{1-6}$$

1.3 *Practical considerations*

1.3.1 *Voltage and pressure levels*

Voltage and pressure levels (V_J, V_W, S_M and S_E) as used in the definitions above may be measured using exactly the same procedure as used in measuring the corresponding levels (i.e. V_J, E_J, P_M, P_E) in Recommendation P.64.

1.3.2 *Analysis bandwith*

The loudness equation given above in Equation 1-5 is broadly applicable to any arbitrary bandwidth. However, for most transmission planning purposes the bandwidth generally selected is 300-3300 Hz. This is because the use of partial connection ratings as engineering tools implicitly requires that for any given connection, the sum of the partial ratings (for example, transmitting plus receiving) should approximately equal the overall rating. Thus the bandwidth used to obtain these ratings should approximate the bandwidth of the most restrictive element(s) in order to avoid cumulating bandwidth penalties when summing partial ratings. The specific limits of 300 Hz and 3300 Hz were selected largely on the basis of bandwidth capabilities of broad-band carrier systems with a 4 kHz channel spacing. In some cases, for example evaluation of a telephone sidetone path, a wider analysis band (e.g. 100-5000 Hz) may permit better estimation of the loudness loss. The method described above may still be used in such cases.

It should be noted that if an actual reference system is constructed for subjective comparison purposes, the system response at 300 and 3300 Hz shall be down 3 ± 1 dB relative to the midband response. The gain of the system shall be adjusted to compensate for the finite slope of the filter skirts (i.e. in comparison to the infinite slope inherent in the definition of § 1.2.2) and deviation from flatness of the pass-band. The amount of this adjustment can be determined by first calculating the OLR (§ 1.2.3) over a frequency range that includes at least the −50 dB points of the real response, and next calculating the OLR of the ideal response over the same frequency range. The difference between the OLRs is the required gain adjustment.

1.3.3 *Number of frequency points*

As a practical matter, measurement frequencies from which a loudness computation is made may be evenly spaced on either a linear frequency scale (1) or logarithmic frequency scale (2). For (1), no fewer than 31 frequencies should be used. For (2), no fewer than 12 frequencies should be used, but there is no significant improvement in accuracy if more than 20 frequencies are used.

1.3.4 *Conversion factors between IEEE and Rec. P.79 loudness ratings*

The following empirical conversion factors have been found useful among North American Administrations for converting between loudness ratings derived according to the IEEE method described above and loudness ratings derived according to Rec. P.79, for 500-type (or equivalent) telephones using the G-handset.

Send: SLR (P.79) = TOLR (IEEE) + 56 dB

Receive: RLR (P.79) = ROLR (IEEE) − 50 dB

Overall: OLR (P.79) = OOLR (IEEE) + 6 dB

Sidetone: STMR (P.79) = SOLR + 8 dB

For send, receive and overall, these relationships give agreement between the different ratings with a tolerance of about ± 1 dB; for sidetone the tolerance is about ± 2 dB.

1.4 *Conclusions*

An alternative algorithm for calculating loudness ratings has been described. This algorithm has been in widespread use in North America for several years and has been found very satisfactory both for transmission planning purposes and characterization of individual network components. One of the main advantages is its relative simplicity.

2 **Algorithms for calculation of loudness ratings** (Contribution from the Australian Administration)

2.1 *Introduction*

There is growing evidence (see § 4) that the algorithm defined in Recommendation P.79 for the calculation of loudness ratings (LRs) is non-optimum, giving undue weight to the lower frequencies. This prompted a study within Telecom Australia to seek a better algorithm. The approach involved determining the loudness rating of many telephone paths and then optimizing the parameters in the algorithm for best agreement between subjective and computed values.

An insert earphone type headset and a (pseudo) loudspeaking telephone were also included in the programme of work. In view of the physical differences from handset telephones, particularly on receiving, it was expected that different algorithms would be required.

2.2 *Basic algorithm*

A method for the computation of loudness ratings (LRs) is derived in Recommendation P.79 and results in a formula of the form:

$$LR = -10/m \log \Sigma\, 10^{(S_i - W_{oi})m/10}$$

where:

- m is the loudness growth coefficient
- S_i is the overall acoustic-acoustic sensitivity in dB of the unknown telephone path (completed by the IRS, if necessary)
- W_{oi} is the (negative) weighting function of frequency, in dB
- i is the 1/3 octave (strictly 1/10 decade) frequency step number.

In the derivation, S_i refers to real mouth and real ear sensitivities, but if the correction factors for using artificial equivalents are included in the definition of W_{oi}, then S_i can be re-defined to be the measured sensitivity with artificial mouth and ear. W_{oi} also includes other components such as the spectral density of human speech, the frequency sensitivity of the human ear, and normalization so that computed loudness rating of the IRS + IRS connection is 0 dB.

2.3 *Determination of parameters*

The weighting function in Rec. P.79 was derived by determining each of the above components and then combining them. In the present work, the weighting function was derived directly. This direct approach leads naturally to consideration of non-handset telephones, such as headsets which may have insert type receivers and handsfree loudspeaking telephones. In the latter case the weighting function must also take into account the diffraction of sound around the human head, the effect of listening with two ears instead of one, and the use of an open rather than occluded ear.

The method involved the insertion of a series of five low-pass and five high-pass filters into various telephone connections, measuring the LR of each subjectively, and then optimizing the parameters to give best agreement (in a least-squares error sense) with the computed values. The overall acoustic-acoustic sensitivities of each connection were first measured using an artificial mouth (B&K type 4219) and an artificial ear (IEC type 318 by B&K) for handsets, and IEC type 711 (B&K type 4157) for the insert receiver.

2.4 *Telephone paths*

The telephone paths involved several different telephone types which are in use in Australia, and are listed in the first column of Tables 2-1 to 2-5. If necessary, the connection was completed using the appropriate IRS end. Since the 802 type was fitted with a carbon transmitter, the send and receive sensitivities were measured using a speech weighted random noise signal. The pseudo loudspeaking telephone (LST) paths were similarly measured to reduce the effect of standing waves in the test room. All other telephones were measured using sine waves. The equalized IRS connections were obtained by first equalizing to give a reasonably flat overall sensitivity (measured objectively) and then adding further equalizers to give either a falling response or a rising response (about 6 dB/octave in both cases). The Featherset headset has an insert type receiver and a noise cancelling electret microphone which is held near the side of the mouth by a boom.

The pseudo loudspeaking telephone for send measurements consisted of a 1/2 inch condenser microphone plus measuring amplifier with a sound level meter A – weighting function. The microphone was mounted on a goose-neck extension piece which held the microphone just above the surface of the table. For receive measurements, the equipment consisted of a power amplifier and a small loudspeaker lying on the long side of its enclosure, with the axis horizontal and pointing to the listener. A real loudspeaking telephone was not used to avoid complications associated with voice switching.

2.5 *Form of weighting function*

Various parametric forms of the weighting function were tried, but a parabola gave almost as good a result as more complicated forms, including higher order polynomials. A parabola can be described in terms of the coordinates of its minimum (in this case) and a coefficient controlling its breadth, by a procedure known as "completing the square", viz.

$$W_{oi} = A + C(i - B)^2$$

In order to compare the weighting functions derived using different values of loudness growth coefficients m, it is more meaningful to consider the product $W_{oi}m$. This quantity may be interpreted as being proportional to the negative of the decibel equivalent of the weighting function which multiplies the band loudness (as distinct from band power) in each of the 1/3 octave (1/10 decade) frequency bands.

The value of i ranges from 0 to 17 for frequencies from 100 Hz to 5012 Hz.

2.6 *Optimum parameters*

The optimum values of m, Am, B, C and Cm are given in Table 2-1 for the various telephone paths considered. Also included in the table are the subjective-objetive error standard deviations (means = 0 dB) and the computed LR of the IRS + IRS connection (which ideally should be 0 dB).

The standard deviations range from 0.1 to 0.4 dB, showing good fit of the model whem optimized for the particular path. Examination of the distribution of the individual errors showed no trends with filter cut-off frequencies. The values of B, Cm and m are fairly consistent with different paths, the biggest differences in B

occurring with the different equalizer responses used with the IRS. Note that although the Featherset and loudspeaking telephone have quite different receive characteristics, *B*, *Cm* and *m* are within the range of those for conventional handset telephones. *Am* is significantly different, however, and this is reflected in the error of the computed LR of the IRS + IRS connection. This suggests that a single frequency weighting shape may be satisfactory for all telephones, whether handset, headset or handsfree, provided that a constant correction factor is applied in certain cases.

Note that the value *A* (and hence *Am*) for the loudspeaking telephone on receive is now believed to be in error. This is discussed later in § 2.11.

TABLE 2-1

Optimum parameters for each path and error statistics

Path	Parameters						Errors	
	m	A	Am	B	C	Cm	Std. dev.	IRS
802 send	0.255	39.67	10.12	9.64	1.225	0.312	0.2	−0.6
802 receive	0.249	42.72	10.63	9.04	0.889	0.221	0.3	0.1
Flip-phone send	0.308	34.72	10.69	9.35	0.732	0.226	0.3	−1.6
Flip-phone receive	0.286	40.69	11.63	8.85	0.513	0.147	0.4	0.5
807 send	0.315	36.38	11.46	9.66	0.648	0.204	0.2	−0.3
807 receive	0.263	43.37	11.41	8.95	0.533	0.140	0.1	0.1
Commander T210 send	0.312	33.84	10.56	9.45	0.934	0.291	0.5	−0.9
Commander T210 receive	0.279	38.28	10.68	8.72	0.704	0.196	0.4	−0.9
Siemens Trans. Cour. send	0.290	35.83	10.39	9.50	1.119	0.325	0.4	−0.2
Siemens Trans. Cour. receive	0.337	35.69	12.03	9.33	0.751	0.253	0.3	2.3
Equalized, IRS flat	0.270	42.47	11.47	9.64	0.581	0.157	0.3	−0.1
Equalized, IRS falling	0.299	40.21	12.02	10.31	0.398	0.119	0.2	−0.6
Equalized, IRS rising	0.300	35.07	10.52	6.66	0.496	0.149	0.3	0.3
Featherset send	0.285	36.48	10.40	9.55	0.684	0.195	0.3	−3.2
Featherset receive	0.330	42.63	14.07	9.28	0.525	0.173	0.3	6.9
Pseudo LST send	0.244	40.29	9.83	8.89	0.776	0.189	0.4	−3.8
Pseudo LST receive	0.232	27.36	6.35	9.40	0.352	0.082	0.3	−23.4

2.7 *Global optimization*

Parameters *A*, *B* and *C* are partly dependent on loudness rating specifics, but *m* is a pure psycho-acoustic phenomenon. The average value of *m* in Table 2-1 is 0.2855 (median = 0.29). The optimization process was therefore repeated with *m* held at 0.2855, with results given in Table 2-2. The standard deviations increased only slightly (about 0.1 dB) verifying that a single value of *m* is practicable.

TABLE 2-2

Optimum parameters and error statistics with m = 0.2855

Path	Parameters					Errors	
	A	Am	B	C	Cm	Std. dev.	IRS
802 send	39.91	10.25	9.64	1.208	0.345	0.3	−0.4
802 receive	37.68	10.76	9.08	0.901	0.257	0.4	0.3
Flip-phone send	37.31	10.65	9.29	0.712	0.203	0.3	−1.7
Flip-phone receive	40.70	11.62	8.85	0.513	0.147	0.4	0.5
807 send	39.76	11.35	9.57	0.603	0.172	0.3	−0.5
807 receive	40.17	11.47	9.03	0.569	0.162	0.2	0.3
Commander T210 send	36.55	10.44	9.38	0.919	0.262	0.5	−1.1
Commander T210 receive	37.46	10.69	8.74	0.711	0.203	0.4	−0.8
Siemens Trans. Cour. send	36.42	10.40	9.49	1.111	0.317	0.4	−0.2
Siemens Trans. Cour. receive	41.05	11.72	9.21	0.691	0.197	0.5	1.9
Equalized, IRS flat	40.16	11.47	9.63	0.606	0.173	0.3	−0.1
Equalized, IRS falling	42.03	12.00	10.43	0.373	0.107	0.3	−0.6
Equalized, IRS rising	36.92	10.54	6.56	0.476	0.136	0.3	0.3
Featherset send	36.42	10.40	9.55	0.685	0.196	0.3	−3.1
Featherset receive	47.08	13.44	9.06	0.490	0.140	0.4	6.4
Pseudo LST send	34.42	9.83	9.02	0.817	0.233	0.5	−3.4
Pseudo LST receive	18.75	5.35	9.28	0.390	0.111	0.4	−23.1

Next, parameters m, B and C were optimized globally, but individual values of A were permitted, to investigate the feasibility of using the same shape for the weighting function, for all telephone types (handset, headset and handsfree), but permitting a correction constant if necessary. Optimization gave $m = 0.2855$, $B = 9.19$ and $C = 0.7723$, with A and errors as shown in Table 2-3. The standard deviations have now increased significantly, the worst being for the IRS with rising frequency response. The errors for this path also show a clear trend with filter cut-off frequency, indicating a lack of fit of the model. Note however that the standard deviations for the headset and handsfree telephone are still comparable with handset telephones in general. The value of A necessary to give a computed LR of 0 dB for the IRS is 38.45.

Table 2-4 gives the errors for a new algorithm (denoted D4 for convenience) based on the above data. The most significant mean errors are −22.4 dB (but see § 2.11) for the loudspeaking telephone receive, 6.9 dB for the headset receive, −3.6 dB for loudspeaking telephone send and −3.0 dB for headset send. There are obvious reasons why the mean errors on receive would not be zero, but the main reason for the errors on send are thought to be due to incorrect pressure distribution as a function of distance of the artificial mouth (B&K type 4219). Another reason might be due to the handset mouth cap affecting the pressure of the feedback microphone in the artificial mouth, while no significant effect occurred with the headset and loudspeaking telephone. Errors for handset telephones are smaller but unfortunately not negligible. These are thought to be mainly due to limitations of the artificial mouth and ear, including the effect of earcap leakage which is not modelled at all, and has to be included in the weighting function.

TABLE 2-3

Optimum A and errors for the case of other parameters globally optimized

Path	A	High pass (Hz)					Low pass (Hz)					Errors	
		158	225	380	630	1020	630	780	1260	2040	3120	Std. dev.	IRS
802 send	38.03	-0.3	-0.6	-0.2	-0.2	-0.6	2.2	0.7	-0.2	-0.4	-0.5	0.9	-0.4
802 receive	38.62	-0.6	0.1	0.3	0.6	0.9	0.0	-0.8	-0.1	-0.5	-0.1	0.5	0.2
Flip-phone send	36.82	0.0	0.0	-0.2	0.0	-0.8	0.3	0.2	0.4	-0.3	0.5	0.4	-1.6
Flip-phone receive	38.95	0.3	0.0	-0.3	0.9	-0.3	-0.3	-0.5	0.5	0.5	-0.7	0.5	0.5
807 send	38.42	-0.4	-0.2	-0.3	-0.5	-1.9	1.2	1.1	0.6	0.5	0.1	0.9	-0.1
807 receive	38.84	-0.2	-0.2	0.1	0.2	0.2	-0.3	0.1	0.5	-0.1	-0.5	0.3	0.4
Commander T210 send	37.27	-0.3	0.2	0.5	0.7	-1.0	0.8	-0.1	-0.3	-0.2	-0.3	0.6	-1.2
Commander T210 receive	37.30	0.0	0.4	0.2	1.3	0.9	-1.0	-1.4	-0.1	-0.3	-0.1	0.8	-1.2
Siemens Trans. Cour. send	38.25	-0.1	-0.3	-0.1	0.4	-0.9	1.5	0.4	-0.6	-0.1	0.0	0.7	-0.2
Siemens Trans. Cour. receive	40.53	1.0	0.0	0.1	-0.2	-0.8	-0.3	0.0	0.1	-0.2	0.2	0.5	2.1
Equalized, IRS flat	38.89	-0.5	0.0	-0.9	-0.7	-1.1	1.0	0.9	1.2	0.5	-0.4	0.8	0.4
Equalized, IRS falling	39.37	0.1	0.1	-0.1	-0.1	-2.2	0.1	1.0	0.8	0.6	-0.3	0.9	0.9
Equalized, IRS rising	37.39	0.8	1.2	2.1	3.8	5.4	-7.1	-4.2	-1.8	0.0	-0.2	3.7	-1.1
Featherset send	35.47	-0.3	-0.3	0.0	-0.4	-1.4	1.6	0.8	0.6	0.4	-1.0	0.9	-3.0
Featherset receive	45.31	0.2	0.0	0.2	0.3	-1.5	-0.1	0.8	0.3	0.2	-0.3	0.6	6.9
Pseudo LST send	34.81	-0.2	-0.3	0.3	0.9	0.9	-0.1	-0.7	0.2	-0.8	0.0	0.6	-3.6
Pseudo LST receive	16.07	-0.4	-0.5	-0.6	0.3	0.0	-0.4	0.9	0.9	0.4	-0.4	0.6	-22.3

TABLE 2-4

Errors for algorithm D4

Path	High pass (Hz)					Low pass (Hz)					Errors	
	158	225	380	630	1020	630	780	1260	2040	3120	Mean	Std. dev.
802 send	-0.7	-1.0	-0.6	-0.7	-1.0	1.8	0.3	-0.7	-0.8	-0.9	-0.4	0.9
802 receive	-0.4	0.2	0.5	0.7	1.1	0.2	-0.6	0.1	-0.3	0.1	0.2	0.5
Flip-phone send	-1.6	-1.7	-1.9	-1.6	-2.4	-1.4	-1.4	-1.2	-1.9	-1.1	-1.6	0.4
Flip-phone receive	0.8	0.5	0.2	1.4	0.2	0.2	0.0	1.0	1.0	-0.2	0.5	0.5
807 send	-0.5	-0.3	-0.4	-0.5	-2.0	1.1	1.0	0.5	0.4	0.1	-0.1	0.9
807 receive	0.2	0.2	0.5	0.6	0.6	0.1	0.5	0.9	0.3	-0.1	0.4	0.3
Commander T210 send	-1.4	-0.9	-0.7	-0.5	-2.1	-0.4	-1.3	-1.5	-1.4	-1.4	-1.2	0.5
Commander T210 receive	-1.1	-0.8	-1.0	0.2	-0.2	-2.1	-2.5	-1.2	-1.5	-1.2	-1.1	0.8
Siemens Trans. Cour. send	-0.3	-0.5	-0.3	0.2	-1.1	1.3	0.2	-0.8	-0.3	-0.2	-0.2	0.7
Siemens Trans. Cour. receive	3.1	2.1	2.1	1.9	1.2	1.8	2.1	2.2	1.9	2.3	2.1	0.5
Equalized, IRS flat	0.0	0.4	-0.4	-0.2	-0.6	1.4	1.4	1.7	1.0	0.0	0.5	0.8
Equalized, IRS falling	1.0	1.0	0.8	0.8	-1.3	1.1	1.9	1.7	1.6	0.6	0.9	0.9
Equalized, IRS rising	-0.3	0.1	1.1	2.7	4.3	-8.2	-5.2	-2.8	-1.1	-1.3	-1.1	3.7
Featherset send	-3.2	-3.3	-3.0	-3.4	-4.4	-1.4	-2.1	-2.4	-2.6	-4.0	-3.0	0.9
Featherset receive	7.0	6.9	7.0	7.1	5.4	6.7	7.7	7.1	7.1	6.6	6.9	0.6
Pseudo LST send	-3.9	-3.9	-3.3	-2.8	-2.7	-3.8	-4.4	-3.4	-4.4	-3.7	-3.6	0.6
Pseudo LST receive	-22.8	-22.9	-23.0	-22.1	-22.3	-22.8	-21.5	-21.5	-22.0	-22.8	-22.4	0.6

2.8 *Comparison with other algorithms*

Table 2-5 compares algorithm D4 with other algorithms. D2 is an algorithm based on preliminary work in which $m = 0.2976$ and the weighting function is defined by $A = 40.50$, $B = 9.867$ and $C = 0.423$. P.79 is the current Recommendation while P.XXE is the draft upon which it is based. Note that draft Rec. P.XXE as published in [2] is in error. On page 178 it states that the mean L_{RME} is -4.72 dB, but in fact it should be -0.1 dB. Thus 4.6 dB should be subtracted if the tabulated data are used. Zw is a complicated algorithm based on the work of E. Zwicker and published as ISO Rec. R532B.

The paths are as previously discussed except that the first item is the set of 14 sidetone responses reported in an earlier work [3].

The group mean errors for the handset telephones are fairly small for all algorithms, but the group standard deviation for P.79 is about twice that of the others. Note that the complicated Zw method does not seem to offer any significant advantage, and still gives rather large errors for the IRS + IRS + equalizer connections. Naturally D4 gives a reasonably good fit because it was optimized for these conditions.

The values of $W_{0i}m$ as a function of frequency for algorithms P.XXE, P.79, D2 and D4 are shown in Figure 2-1. Note that P.XXE and D4 are very similar, and that P.79 shows much smaller (negative) weight at low frequencies.

TABLE 2-5

Comparison of errors for five algorithms

Path	D2		D4		P.79		P.XXE		Zw	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Sidetone [3]	-1.2	0.5	-1.2	0.7	1.9	1.4	-0.8	0.6	-0.6	0.7
802 send	-0.7	0.9	-0.4	0.8	-1.4	3.4	-1.0	1.2	-0.4	1.8
802 receive	-0.4	1.5	0.2	0.5	0.1	2.2	-0.1	0.5	0.6	0.9
Flip-phone send	-2.2	1.0	-1.6	0.4	-1.9	3.0	-1.9	0.7	-1.1	1.7
Flip-phone receive	0.1	1.6	0.5	0.5	-0.5	2.6	0.2	0.5	1.3	1.3
807 send	-0.6	0.5	0.0	0.9	-0.5	3.6	-0.3	1.1	0.7	2.1
807 receive	-0.2	1.2	0.4	0.3	-0.3	2.6	0.0	0.3	1.0	1.4
Commander T210 send	-1.6	1.4	-1.2	0.5	-2.7	2.5	-1.6	0.8	-0.6	1.7
Commander T210 receive	-2.1	1.7	-1.2	0.8	-1.3	1.8	-1.5	0.6	-0.7	1.0
Siemens Trans. Cour. send	-0.9	1.0	-0.2	0.6	-0.6	3.1	-0.6	1.0	0.0	2.1
Siemens Trans. Cour. receive	1.1	1.0	2.1	0.4	2.1	2.7	1.8	0.8	2.8	1.3
Equalized, IRS flat	0.0	0.4	0.4	0.8	3.2	3.6	0.5	0.9	1.4	2.1
Equalized, IRS falling	0.1	0.5	0.9	0.8	4.7	3.8	0.9	1.2	2.1	2.2
Equalized, IRS rising	-1.1	4.7	-1.1	3.5	0.3	1.5	-1.0	3.4	-0.5	2.5
Featherset send	-3.3	0.8	-3.0	0.8	-4.1	3.1	-3.3	1.0	-2.4	2.0
Featherset receive	6.2	1.1	6.9	0.6	5.0	2.5	6.3	0.8	7.2	1.6
Pseudo LST send	-4.4	1.6	-3.6	0.6	-4.3	1.8	-4.0	0.5	-3.3	0.9
Pseudo LST receive	-23.4	0.7	-22.4	0.5	-22.0	2.5	-22.7	0.6	-21.9	1.0
Handset	-0.61	0.95	-0.09	1.02	-0.09	2.09	-0.35	1.06	0.51	1.19

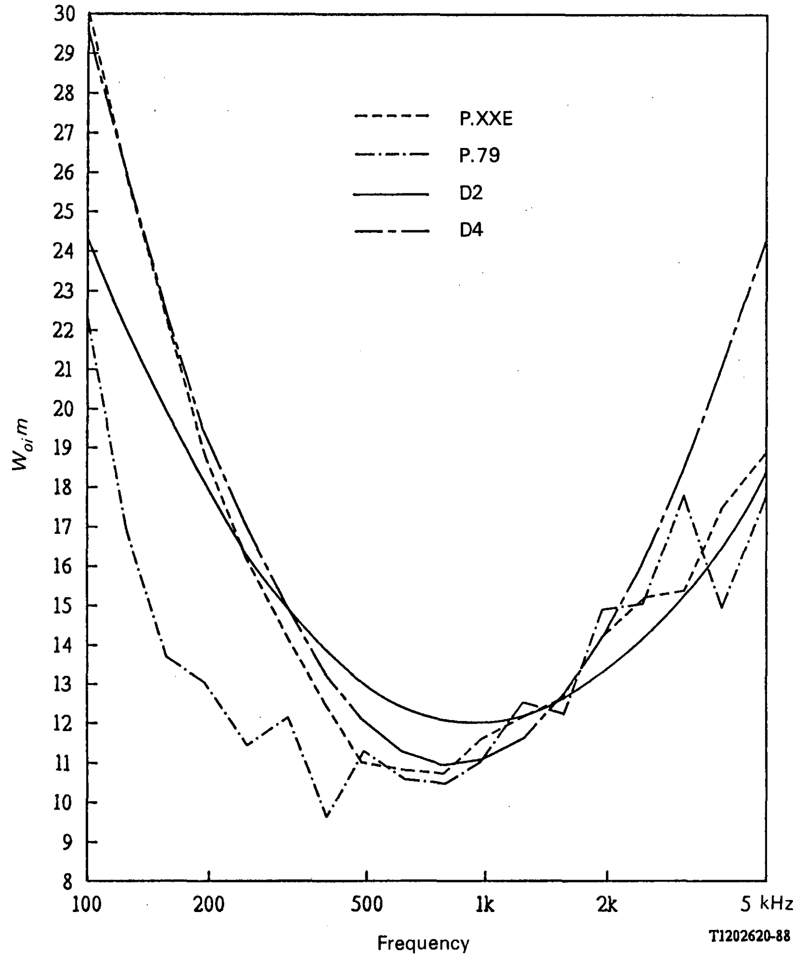


FIGURE 2-1

Loudness weighting functions $W_{0i}m$ as a function of frequency

2.9 *Validation of new algorithm*

Table 2-6 gives the subjective-objective errors for test results which are not used in deriving D4. Two samples of 802 telephone (local designations 82/YZ and 82/IA) were each fitted with one of four 20E non-carbon transmitters (designated 101, 165, 310 and 313) for send measurements. Only one receive measurement was made for each telephone. Three lines were used, viz. zero, 1.6 km and 4.2 km of 0.4 mm cable.

One consistent trend is that the errors become more positive with increasing line length, and range from 0.6 dB for D2 through 0.9 dB for D4, P.XXE and Zw, to 1.2 dB for P.79. A possible reason for this trend is the progressive high frequency loss which occurs with line length, and inadequacies in the loudness models to cope with this. This is also consistent with the errors associated with the equalized IRS results in Table 2-5, where the falling response gives the most positive error and the rising response the most negative error of the set of three.

TABLE 2-6

Errors for 802 telephones (20E non-carbon transmitters) plus lines for five algorithms

Algorithm	Line (km)	82/YA					82/IA					Mean	Std. dev.
		Send				Receive	Send				Receive		
		Transmitter					*Transmitter*						
		101	165	310	313		101	165	310	313			
D2	Zero	0.2	0.0	-0.4	-0.5	-0.1	0.1	-0.1	-0.9	-0.1	-0.5	-0.2	0.3
	1.6	-0.2	0.1	0.3	-0.1	-0.3	0.8	-0.3	-0.1	0.6	0.5	0.1	0.4
	4.2	0.8	0.7	0.6	0.5	0.2	0.4	-0.1	0.1	0.6	0.0	0.4	0.3
D4	Zero	0.8	0.6	0.2	0.0	-0.3	0.7	0.6	-0.2	0.5	-0.6	0.2	0.5
	1.6	0.5	0.8	1.0	0.5	-0.4	1.5	0.4	0.7	1.2	0.5	0.7	0.5
	4.2	1.7	1.6	1.5	1.3	0.5	1.3	0.9	1.1	1.4	0.2	1.1	0.5
P.79	Zero	1.1	1.0	0.5	0.2	0.1	1.1	0.9	0.2	0.8	-0.2	0.6	0.4
	1.6	0.9	1.2	1.4	0.8	0.2	1.9	0.8	1.1	1.5	1.1	1.1	0.4
	4.2	2.3	2.2	2.1	1.7	1.3	1.9	1.5	1.7	2.0	1.1	1.8	0.4
P.XXE	Zero	0.5	0.4	-0.1	-0.4	-0.3	0.4	0.3	-0.6	0.1	-0.6	0.0	0.4
	1.6	0.2	0.5	0.6	0.1	-0.4	1.2	0.2	0.3	0.8	0.5	0.4	0.4
	4.2	1.4	1.3	1.1	0.9	0.4	1.0	0.6	0.7	1.1	0.2	0.9	0.4
Zw	Zero	0.2	0.1	-0.4	-0.7	0.0	0.1	0.1	-0.8	-0.2	0.0	-0.2	0.3
	1.6	0.0	0.3	0.3	-0.2	-0.2	1.0	-0.1	0.0	0.5	1.0	0.3	0.4
	4.2	1.2	1.1	0.9	0.6	0.5	0.8	0.4	0.5	0.8	0.5	0.7	0.3

2.10 *Attempts to reduce errors*

In order to explore whether another weighting function would simultaneously give small errors for the 802 telephone only, with both filters and lines, the 802 + lines data described above was combined with the 802 + filter data described earlier. A new weighting function was then optimized, with A constrained to give 0 dB error for the LR of the IRS. It was found however that the optimum parameters were not greatly different from those in D4 and that the range of errors with line length was only reduced by 0.1 dB to 0.8 dB.

It was thought possible that forcing a polynomial fit to the weighting function may be partly responsible for this poor agreement, so a piecewise linear weighting function was tried, with break frequencies at i = 4, 7, 10 and 13 (f = 250, 500, 1000 and 2000 Hz respectively). It was found that the range of errors with line length was unchanged at 0.8 dB. Thus the weighting function shape does not seem to be at fault.

A simplification inherent in all algorithms from P.XXE to D4 is that the weighting function does not cause any frequency band to be masked, whereas it is assumed in the derivation of these models that it is only the band loudness above threshold which contributes to loudness. The basic formula was therefore changed to include a threshold rather than a weighting function. Summation is only over those bands which are above threshold. A disadvantage of this algorithm is that it is now not possible to make loudness rating the subject of the formula, and an iterative approach is necessary. A parabolic threshold function was assumed, and it was found that the range of errors with line length was only reduced a further 0.1 dB to 0.7. The marginal improvement does not justify the extra complication of this method.

Finally, the effect of frequency masking was included by investigating whether a better way of using Zwicker's loudness algorithm could be found. In addition to the sensitivity of hearing which is inherent in Zwicker's algorithm, a LR algorithm must also include the spectral density and level of the speech signal, the ear cap leakage loss and the junction loss to give the same loudness through the IRS + IRS path as the NOSFER system with 25 dB in its junction. These may be combined to form an auxiliary function analogous to an input signal to the telephone path, where the output is fed to Zwicker's loudness algorithm. Assuming a parabolic shape to this auxiliary function, it was found that the range of mean errors with line length was 0.8 dB and thus comparable to that of previous algorithms, such as D4.

A possible reason why none of the methods was successful in reducing errors to a low and random value (i.e. no trend with line length) may be that the subjects changed their bases of listening to the speech from one filter condition to the next. They may not listen to the signal as a whole, but base their comparison on a smaller band or bands where the main energy lies (formants). The location of the band or bands could vary depending on the cut-off frequencies of the filters. Zwicker based his method on subjective data gathered on non-speech signals, but it is known that people listen to speech in a different way to other sounds, and this may affect the judgement of loudness. Other possible sources of discrepancy are possible, including the effect of changes in the voice-ear team membership during the course of the investigation.

2.11 *Postscript on the correction factor for loudspeaking telephone receive*

The receive correction factor found initially for the loudspeaker and amplifier combination was about −22.4 dB, but in subsequent work a drift in this value was observed. Whether this was due to set-up errors, hardware faults or to changing bases of rating loudness by the voice-ear team has not been resolved. Subsequent tests repeating those reported above and others have yielded a correction factor of about −14.0 dB, and this is now believed to be more correct. (The D4 loudness algorithm continued to give good consistency in the repeat tests, with a standard deviation of 0.7 dB over the range of filters.)

2.12 *Conclusion*

A revised algorithm has been found which is remarkably similar to the draft Recommendation upon which the present Recommendation P.79 was based. Using either of these methods gives about half the standard deviation of the difference between subjective and objective measurements which would be obtained with Rec. P.79. A general accuracy of about ± 2 dB can be expected, which is about the order of accuracy of subjective tests, but with better repeatability and lower cost.

Although it was expected that a different weighting function would be required for headsets and loudspeaking telephones, in fact it was found that a constant correction for each path type proved to be all that was necessay for practical purposes. In particular, the following corrections should be added to the calculated LRs:

Headset

Send: −3.0 dB
Receive: 6.9 dB (insert receiver only)

Loudspeaking telephone

Send: − 3.6 dB
Receive: −14.0 dB

As far as the revision of Rec. P.79 is concerned, two courses of action seem possible. Preferably,

i) pool all the data available worldwide and derive a global average using the principles described above,

or alternatively

ii) return to the algorithm weights of draft Rec. P.XXE.

3 **Uniform algorithms for the calculation of R25 equivalents and loudness ratings** (from the Ministry of Post and Telecommunications of the People's Republic of China)

3.1 *Introduction*

The subjective test team of the CCITT Laboratory has been changed since 1985. From the periodic stability check reports of the CCITT Laboratory, it can be ascertained that the recent subjectively determined value x_2 (see Recommendation P.78) is about 18 dB which is close to the value determined at other laboratories, and different from the previously determined value of 12 dB. In addition, the SR25E and RR25E values of telephone systems determined recently by the CCITT Laboratory are several decibels lower than the results previously obtained, and close to those measured by other laboratories.

In this connection, it is possible to use a uniform algorithm, similar to the simple algorithm in Recommendation P.79, for the calculation of R25 equivalents and loudness ratings, with values for the slope parameter m and the G-functions different from those given in Recommendation P.79.

In order to obtain a suitable algorithm and appropriate parameters, four different algorithms were used in order to calculate the values of R25E and LR, and the results were compared. Three of them are similar to that used for the calculation of loudness ratings described in Recommendation P.79, except that different values of the slope parameter m and the G-functions are used.

These values:

- are taken from draft Recommendation P.XXE [2];
- correspond to the Chinese test team;
- correspond to the old test team of the CCITT Laboratory, but with L_E corrected in the NOSFER receiving system.

The fourth algorithm used was the ISO-532B (Zwicker) algorithm.

3.2 *Comparison of various algorithms*

The four algorithms used here are labelled as the P.XXE, the Chinese, the P.79 Cor. and the ISO-532B algorithms.

3.2.1 *SFC of the reference system*

3.2.1.1 The sensitivity/frequency characteristic (SFC) data of the sending system and the receiving system (without leakage) of the NOSFER are taken from Recommendation P.42 (Red Book). The coupling loss at the receiving part of the NOSFER is included in the receiving SFC in the calculation.

Several years ago the Chinese Administration pointed out that the SFC data of the NOSFER receiving system measured by the IEC 318 artificial ear with the flat plate differed considerably from those measured with the operator's ear, and measured the values of L_E corresponding to the earphone type DR-701 used by the Chinese test crew in the receiving system of the NOSFER.

This point of view has been verified by many Administrations and has been generally accepted by CCITT Study Group XII. The values of L_E used here are those corresponding to the CCITT Laboratory test team, as given by the French Administration (Contribution COM XII-111, 1985-1988) (see Table 3-1).

3.2.1.2 The SFC data of IRS are taken from Recommendation P.48 and the SFC values of the receiving system are corrected using the L_E given in Recommendation P.79.

3.2.2 *Slope parameter m and G-functions*

Methods for estimating m and G are described in Contribution COM XII-3 and COM XII-10 (1981-1984).

3.2.2.1 *P.79 Cor. algorithm (m = 0.175)*

The values of the slope parameter m and the G-functions in Recommendation P.79 are derived from the results of the filter loudness loss test of the old CCITT Laboratory test team; the leakage between the ear of the operator and the earphone of NOSFER is not included. The values of the G-functions given in Recommendation P.79 must therefore be corrected. The results of the G-functions with correction of L_E are listed in Table 3-2.

TABLE 3-1

Acoustic coupling loss L_E used in calculation

Frequency	L_E NOSFER	L_E P.79
100	0.9	20.0
125	0.2	16.5
160	−0.6	12.5
200	−1.6	8.4
250	−2.9	4.9
315	−4.2	1.0
400	−5.3	−0.7
500	−5.4	−2.2
630	−4.9	−2.6
800	−4.6	−3.2
1000	−4.5	−2.3
1250	−3.9	−1.2
1600	−4.6	−0.1
2000	−3.3	3.6
2500	−3.2	7.4
3150	−3.3	6.7
4000	−3.7	8.8
5000	−2.9	10.0
6300	−0.8	12.5
8000	−0.8	15.0

TABLE 3-2

10 $\log_{10}$ G of various algorithms

Frequency	P.79 Cor.	P.XXE	Chinese
100	−31.86	−35.90	−30.67
125	−28.58	−34.11	−30.63
160	−27.14	−32.94	−30.68
200	−28.13	−31.50	−30.81
250	−28.48	−30.96	−31.02
315	−31.22	−31.21	−31.35
400	−30.10	−31.15	−31.79
500	−33.02	−30.97	−32.33
630	−33.46	−32.13	−33.00
800	−34.34	−33.05	−33.83
1000	−35.51	−34.50	−34.74
1250	−37.97	−35.91	−35.78
1600	−38.60	−37.14	−37.10
2000	−41.22	−38.50	−38.46
2500	−41.66	−39.66	−39.96
3150	−45.77	−41.11	−41.70
4000	−43.54	−43.45	−43.68
5000	−47.03	−45.37	−45.71
6300	−48.03		−48.01
8000	−46.32		−50.60

3.2.2.2 *P.XXE algorithm (m = 0.225)*

The values for the slope parameter *m* and the *G*-functions are taken from Table 1, page 185 of COM XII-1 [2]. Also see Table 3-2 of this Supplement.

3.2.2.3 *Chinese algorithm (m = 0.2)*

Results of smoothed *G*-functions are used [see Contribution COM XII-233 (1981-1984)]. Values are also given in Table 3-2.

The coupling loss of the NOSFER earphone is not included in the estimation of the *G*-functions but this has little effect on the smoothed result of the *G*-functions.

3.2.3 *W-weights for the calculation of R25E*

Methods for deriving *W*-weights are described in Contributions COM XII-3 and COM XII-10 (1981-1984).

3.2.3.1 *P.79 Cor. algorithm*

Weights are derived from the SFC data of NOSFER described in § 3.2.1.1 and the data for *m* and *G*-functions given in § 3.2.2.1.

3.2.3.2 *P.XXE algorithm*

W-weights are derived from the SFC data of NOSFER described in § 3.2.1.1 and the data for *m* and *G*-functions given in § 3.2.2.2. In the absence of a complete set of data for the *G*-functions at high and low frequencies, a number of arbitrary values have had to be chosen in this contribution.

3.2.3.3 *Chinese algorithm*

W-weights are derived from the SFC data of NOSFER described in § 3.2.1.1 and the data for *m* and the *G*-functions given in § 3.2.2.3.

The derived *W*-weights of the three algorithms discussed above for the calculation of R25E are listed in Table 3-3.

3.2.4 *W-weights for the calculation of LR*

The methods for the derivation of *W*-weights for the three algorithms are similar to those described in § 3.2.3, except that the SFC data of IRS (with the L_E of P.79) are used instead of the SFC data of NOSFER.

The derived *W*-weights of the three algorithms discussed above for the calculation of LR are listed in Table 3-4.

3.2.5 *Source of data for the SFC of telephone systems and the subjectively determined values of R25E and LR*

In making comparisons between the subjectively determined results and the calculated results, use can only be made of the data relating to telephone sets with subjectively determined values established by the new CCITT test team and the corresponding SFC values.

3.2.5.1 *For SR25E*

There are only six sets of sending SFC data provided by three linear telephone sets under 0/L line conditions (i.e. with or without lines). These data are taken from CCITT Laboratory Technical Report 808 (Temporary Document 84, Working Party XII/1, April 1987); the other set of subjectively determined values is taken from CCITT Laboratory Technical Report 797 (Temporary Document 78, Working Party XII/1, April 1987).

3.2.5.2 *For RR25E*

The subjectively determined values of RR25E of some telephone systems are taken from CCITT Laboratory Technical Report 797 and the corresponding SFC data were given by the Head of the CCITT Laboratory in October 1986.

TABLE 3-3

W-weights for the calculation of R25E

	P.79 Cor.		P.XXE		Chinese	
Frequency	W_S	W_R	W_S	W_R	W_S	W_R
100	109.6	116.6	106.9	113.9	92.2	99.2
125	82.4	89.8	92.1	99.5	84.5	91.9
160	67.2	75.2	81.2	89.2	78.5	86.5
200	66.6	76.3	69.5	79.2	73.4	83.1
250	60.6	72.7	60.4	72.5	67.2	79.3
315	67.6	82.2	54.3	68.9	61.0	75.6
400	53.8	70.2	47.9	64.3	56.6	73.0
500	63.5	80.8	41.3	58.6	52.9	70.2
630	60.5	75.5	42.4	57.4	51.5	66.5
800	60.8	74.3	42.9	56.4	51.6	65.1
1000	62.6	75.1	45.5	58.0	51.9	64.4
1250	70.5	81.9	47.2	58.6	51.9	63.3
1600	67.3	79.9	47.1	59.7	52.3	64.9
2000	78.4	90.1	50.3	62.0	55.8	67.5
2500	74.8	86.6	50.6	62.4	57.9	69.7
3150	93.2	102.1	53.5	62.4	62.3	71.2
4000	76.7	84.6	61.3	69.2	69.2	77.1
5000	88.8	103.3	63.2	77.7	72.2	86.7
6300	84.9	110.0	92.2	117.3	74.9	100.0
8000	80.4	99.1	102.7	121.4	93.8	112.5

$m = 0.175$ $m = 0.225$ $m = 0.2$

3.2.5.3 *For SLR and RLR*

The subjectively determined values are taken from CCITT Laboratory Technical Report 771 (Temporary Document 42, Working Party XII/1, May 1986) and the corresponding SFC data [the sending data measured at LRGP (loudness rating guard-ring position)] were also provided by the CCITT Laboratory.

3.2.6 *Method of calculation*

3.2.6.1 For the P.79 Cor., P.XXE and the Chinese algorithms, the equations used for the calculation of SR25E, RR25E, SLR and RLR are as follows:

$$SR25E = -\frac{10}{m} \log_{10} \sum_{i}^{N} 10^{m(SUMJ - W_S)/10}$$

$$RR25E = -\frac{10}{m} \log_{10} \sum_{i}^{N} 10^{m(SUJE - W_R)/10}$$

$$SLR = -\frac{10}{m} \log_{10} \sum_{i}^{N} 10^{m(SUMJ - W_S)/10}$$

$$RLR = -\frac{10}{m} \log_{10} \sum_{i}^{N} 10^{m(SUJE - W_R)/10}$$

TABLE 3-4

W-weights for the calculation of LR

	P.79 Cor.		P.XXE		Chinese	
Frequency	W_S	W_R	W_S	W_R	W_S	W_R
100	149.3	147.6	150.0	150.0	135.7	134.0
125	111.4	112.2	150.0	150.0	117.2	118.0
160	85.3	87.6	150.0	150.0	100.3	102.5
200	74.4	82.5	82.5	90.6	85.0	93.1
250	61.5	73.6	66.7	78.8	71.9	84.0
315	62.2	79.2	54.0	71.0	59.3	76.3
400	46.0	65.0	45.1	64.1	52.4	71.4
500	54.6	75.0	37.4	57.8	47.6	68.0
630	49.4	70.0	36.1	56.7	44.2	64.8
800	48.2	68.6	35.5	55.9	42.6	63.0
1000	50.6	69.2	38.8	57.4	43.6	62.2
1250	59.0	75.0	41.0	57.0	44.2	60.2
1600	57.3	71.0	42.3	56.0	45.8	59.5
2000	71.5	80.7	48.6	57.8	52.5	61.7
2500	71.8	75.7	52.7	56.6	58.8	62.7
3150	88.5	92.6	54.0	58.1	61.5	65.6
4000	116.7	113.5	106.5	103.3	112.7	109.5
5000	155.9	143.2	150.0	150.0	143.0	130.3
6300	170.0	163.6	150.0	150.0	163.8	157.4
8000	180.0	165.0	150.0	150.0	196.7	181.7

$m = 0.175$ $m = 0.225$ $m = 0.2$

It should be noted that:

- different values of m and W-weights are used for the three different algorithms discussed above;
- the values of W_s for the calculation of SR25E are different from the values used for the calculation of SLR within the same algorithm. The same rule applies to W_R for the calculation of RR25E and RLR.
- S_{UMJ} are the sending SFC values of telephone systems measured at RESP (reference equivalent speaking position) and the RGP for the calculation of SR25E and SLR respectively; and
- the L_E values listed in Recommendation P.79 are used to correct the telephone receiving systems for the calculation of RR25E and RLR.

3.2.6.2 *Method using the ISO-532B algorithm*

a) Input signal: the long-term speech spectrum given in Recommendation P.50 and ⅓ octave data are used.

b) Reference system: *NOSFER sending + 25 dB + NOSFER receiving* for R25E, and *IRS sending + 18 dB + IRS receving* for LR.

c) Tested system: varies depending on the items. For example, for SR25E, sending tested + variable attenuator (X) + NOSFER receiving.

Proceed as follows:

1) Calculate the loudness of the reference system using the ISO-532B algorithm on the basis of the output levels in 1/3 octave bands of the reference system.

2) Calculate the loudness of the tested system using the ISO-532B algorithm on the basis of the output levels in 1/3 octave bands of the tested system, change the attenuation value X of the variable attenuator until the calculated loudness is the same as that in the reference system.

3) Then:

$$R\,25\,E = 25 - X$$

$$LR = 18 - X$$

In calculating R25E and LR with the ISO-532B algorithm, the SFC data of the reference systems and the telephone systems are the same as those used in the other algorithms discussed above.

3.2.7 *Calculated results*

The subjectively determined values of SR25E, RR25E, SLR and RLR, the result calculated by using various algorithms, and the differences between the subjectively determined values and the calculated results are given in Tables 3-5 a) to 3-5 d).

For the sake of comparison, the mean results calculated by the four algorithms are summarized in Table 3-6.

3.3 *Discusion*

Before analyzing the calculated results, it is necessary to bear in mind the effect of the diffraction by the human head and the reverberation of the test room on the sending SFC and NOSFER. As a result of this effect, the difference in the sending SFC between the mouth reference point of the NOSFER system and a point 140 mm in front of the operator's lips is less than 13.46 dB under ideal conditions, i.e. with the virtual sound source 6 mm behind the lips being taken to be the actual human sound source and assuming the sound to be transmitted in a free field. In the Chinese subjective test room, this difference has been mesured with an average correction of 1 to 1.5 dB for each frequency (see contribution COM XII-209 (1985-1988)). This effect has not been included in any of the four algorithms discussed above.

3.3.1 The calculated results of SR25E and RR25E using the P.79 Cor. algorithm are about 1.5 to 2 dB higher than the subjectively determined values. This is understandable because the values of slope parameter m and the G-functions were estimated on the basis of the filter test results of the old CCITT test team.

3.3.2 Both the P.XXE and the Chinese algorithms can be used as the uniform algorithm for the calculation of R25E and LR. The SR25E calculated by the P.XXE algorithm is about 1 dB lower than the subjective result, but after correction for diffraction by the human head and the reverberation of the test chamber as discussed in § 3.3, there may be fairly good agreement between the subjective and objective values of SR25. In view of the fact that some values, at high and low frequencies, of the G-functions and W-weights used in the P.XXE algorithm are chosen arbitrarily and that a correction has to be made to the sending SFC of NOSFER, the Chinese algorithm may be better than the P.XXE algorithm in use.

3.3.3 The results calculated using ISO-532B agree with the corresponding subjective test results. It has been noticed, however, that the standard deviation for the mean values of the differences of SR25, SLR and RLR is larger than that for the other algorithms. Furthermore, the ISO-532B algorithm is much more complicated than the other algorithms. This algorithm would not therefore be the best choice.

3.3.4 The difference of the results of SLR and of RLR calculated with the P.79 Cor. algorithm and with the original P.79 algorithm, respectively, is generally less than 0.1 dB.

3.3.5 It is not advisable to use the P.79 Cor. algorithm to calculate R25E values because of the considerable difference between the calculated values and the subjectively determined values.

The difference between the subjectively determined values of SR25E for a telephone set obtained by the old and by the new test team is about 4 to 6 dB, respectively, while the difference calculated by the P.79 Cor. algorithm and by the Chinese algorithm is about 2 dB.

The value of R25E calculated by the P.79 Cor. algorithm does not agree with the subjectively determined value of the old test team either.

3.4 *Conclusion*

A simpler algorithm such as the Chinese algorithm can be used as the standard algorithm for the calculation of R25E and LR. There is good agreement between the calculated results and the results subjectively determined by the new test team of the CCITT Laboratory.

The statement appearing in some Recommendations to the effect that a simple algorithm cannot be used for the comparison of the loudness of wideband systems should be revised.

4 **Loudness rating coefficients derived from subjective measurements on high-pass (HP) and low-pass (LP) filtered speech** (Contribution from ELLEMTEL, Sweden)

4.1 *Introduction*

The exact shape of the frequency-weighting of the loudness rating (LR) algorithm is not very critical when computing LR values for routine planning evaluations. However, reasonable realistic values of the coefficients are needed for a more detailed analysis of, for instance, attenuation distortion and bandwidth restriction effects.

Loudness rating parameters may be derived from known statistics of the "average" speech power spectrum and the "average" hearing threshold frequency response curves.

An alternative direct way is to make use of subjective listening tests of the influence of variable low-pass and high-pass filters in a NOSFER type circuit. Such measurements have been made many times in the past. This section uses four sets of data of which three originate from STL [4] and one from the People's Republic of China [5].

As is well known, subjective evaluation of loudness has its difficulties. A prime requirement is that the test team must represent "ordinary people" with regard to speech and hearing. Also, the team must be instructed to judge specifically "loudness impression" and not "quality impression" of bandwidth limitation. The CCITT test team seems not to have fulfilled these criterions when performing the measurements for the P.79 algorithm.

4.2 *Derivation of loudness rating coefficients*

For complex noise spectra of time-constant nature the masking effects between frequency bands have to be considered, i.e. the Zwicker algorithm should be used for evaluating loudness. However, it is rather doubtful whether this complex method is really necessary (or even correct) for speech signals. Instead, the simpler conventional "physiological loudness impression" model for interpreting the subjective results will be tried.

The expression for the loudness loss A caused by a filter introduced in the electric part of the transmission path of speech sounds from mouth to ear will be given. To facilitate the mathematical treatment, the usual series summation over the third-octave bands is replaced by a continuous integration over a logarithmic frequency scale.

TABLE 3-5

a) **Comparison of subjective and calculated results using the P.79 Cor. algorithm**

SR25E					
Set	Subjective 1	Subjective 2	Calculated	Difference 1	Difference 2
F6/0	+3.35	+2.45	+4.37	+1.02	+1.92
F6/L	+10.97	+10.93	+13.29	+2.32	+2.36
Q1/0	+5.62	+5.33	+6.52	+0.90	+1.19
Q1/L	+15.72	+14.85	+16.01	+0.29	+1.16
E5/0	+11.23	+11.23	+13.39	+2.16	+2.16
E5/L	+20.85	+20.85	+23.52	+2.67	+2.67
Mean				+1.56 dB	+1.91 dB
Std. dev.				0.87 dB	0.57 dB

RR25E			
Set	Subjective	Calculated	Difference
D3/0	−5.45	−2.37	+3.08
D3/L	+0.15	+2.35	+2.20
E5/0	−4.98	−2.97	+2.01
E5/L	+1.58	+2.49	+0.91
F6/0	−1.55	+1.80	+3.35
F6/L	+7.00	+9.47	+2.47
G5/0	−6.67	−5.59	+1.08
G5/L	−0.23	+0.43	+0.66
Q1/0	−2.35	−1.12	+1.23
Q1/L	+6.48	+6.87	+0.39
R2/0	−8.80	−6.95	+1.85
R2/L	−1.13	+0.43	+1.56
T1/0	−2.98	−0.44	+2.54
T1/L	+1.88	+3.82	+1.94
U1/0	−5.25	−2.54	+2.71
U1/L	−1.03	+0.37	+1.40
V1/0	−8.95	−7.25	+1.70
V1/L	−1.75	+0.09	+1.84
Mean			+1.83 dB
Std. dev.			0.79 dB

SLR			
Set	Subjective	Calculated	Difference
D3/0	+4.45	+3.11	−1.34
D3/L	+5.75	+4.16	−1.59
D4/0	+2.55	+2.28	−0.27
D4/L	+3.65	+3.42	−0.23
F6/0	−0.55	+0.64	+1.19
F6/L	+8.95	+9.26	+0.31
F7/0	+1.10	+0.83	−0.27
F7/L	+10.20	+9.59	−0.61
T1/0	+7.80	+8.40	+0.60
T1/L	+12.00	+12.25	+0.25
T2/0	+10.45	+9.12	−1.33
T2/L	+15.00	+13.09	−1.91
U1/0	+5.55	+5.52	−0.03
U1/L	+10.30	+10.02	−0.28
Mean			−0.39 dB
Std. dev.			0.85 dB

RLR			
Set	Subjective	Calculated	Difference
D3/0	−6.85	−6.31	+0.54
D3/L	−1.45	−1.41	+0.04
D4/0	−9.15	−6.98	+2.17
D4/L	−3.75	−2.08	+1.67
F6/0	−2.45	−2.76	−0.31
F6/L	+5.55	+5.04	−0.51
F7/0	−1.50	−2.88	−1.38
F7/L	+6.05	+4.95	−1.10
T1/0	−5.55	−4.17	+1.38
T1/L	−0.30	+0.12	+0.42
T2/0	−3.75	−4.05	−0.30
T2/L	+2.25	+0.78	−1.47
U1/0	−6.85	−6.65	+0.20
U1/L	−2.30	−3.52	−1.22
Mean			+0.01 dB
Std. dev.			1.11 dB

TABLE 3-5 *(continued)*

b) **Comparison of subjective and calculated results using the P.XXE algorithm**

SR25E

Set	Subjective 1	Subjective 2	Calculated	Difference 1	Difference 2
F6/0	+3.35	+2.45	+0.92	−2.43	−1.53
F6/L	+10.97	+10.93	+10.61	−0.36	−0.32
Q1/0	+5.62	+5.33	+4.33	−1.29	−1.00
Q1/L	+15.72	+14.85	+14.02	−1.70	−0.83
E5/0	+11.23	+11.23	+9.54	−1.69	−1.69
E5/L	+20.85	+20.85	+19.84	−1.01	−1.01
Mean				−1.41 dB	−1.06 dB
Std. dev.				0.64 dB	0.45 dB

RR25E

Set	Subjective	Calculated	Difference
D3/0	−5.45	−4.79	+0.66
D3/L	+0.15	+0.68	+0.53
E5/0	−4.98	−5.87	−0.89
E5/L	+1.58	−0.08	−1.66
F6/0	−1.55	−1.58	−0.03
F6/L	+7.00	+6.69	−0.31
G5/0	−6.67	−7.34	−0.67
G5/L	−0.23	−1.00	−0.77
Q1/0	−2.35	−3.08	−0.73
Q1/L	+6.48	+5.00	−1.48
R2/0	−8.80	−8.87	−0.07
R2/L	−1.13	−1.10	+0.03
T1/0	−2.98	−2.91	+0.07
T1/L	+1.88	+1.91	+0.03
U1/0	−5.25	−5.11	+0.14
U1/L	−1.03	−1.43	−0.40
V1/0	−8.95	−9.35	−0.40
V1/L	−1.75	−1.79	−0.04
Mean			−0.33 dB
Std. dev.			0.60 dB

SLR

Set	Subjective	Calculated	Difference
D3/0	+4.45	+3.81	−0.64
D3/L	+5.75	+4.64	−1.11
D4/0	+2.55	+3.08	+0.53
D4/L	+3.65	+3.95	+0.30
F6/0	−0.55	+0.09	+0.64
F6/L	+8.95	+9.46	+0.51
F7/0	+1.10	+0.29	−0.81
F7/L	+10.20	+9.80	−0.40
T1/0	+7.80	+9.02	+1.22
T1/L	+12.00	+13.38	+1.38
T2/0	+10.45	+9.70	−0.75
T2/L	+15.00	+14.13	−0.87
U1/0	+5.55	+6.16	+0.61
U1/L	+10.30	+11.32	+1.02
Mean			+0.12 dB
Std. dev.			0.82 dB

RLR

Set	Subjective	Calculated	Difference
D3/0	−6.85	−5.61	+1.24
D3/L	−1.45	−0.01	+1.44
D4/0	−9.15	−6.22	+2.93
D4/L	−3.75	−0.53	+3.22
F6/0	−2.45	−2.77	−0.32
F6/L	+5.55	+5.69	+0.14
F7/0	−1.50	−2.86	−1.36
F7/L	+6.05	+5.54	−0.51
T1/0	−5.55	−3.63	+1.92
T1/L	−0.30	+1.25	+1.55
T2/0	−3.75	−3.42	+0.33
T2/L	+2.25	+2.01	−0.24
U1/0	−6.85	−6.06	+0.79
U1/L	−2.30	−2.18	+0.12
Mean			+0.80 dB
Std. dev.			1.27 dB

TABLE 3-5 *(continued)*

c) **Comparison of subjective and calculated results using the Chinese algorithm**

SR25E					
Set	Subjective 1	Subjective 2	Calculated	Difference 1	Difference 2
F6/0	+3.35	+2.45	+1.78	−1.57	−0.67
F6/L	+10.97	+10.93	+11.52	+0.55	+0.59
Q1/0	+5.62	+5.33	+5.00	−0.62	−0.33
Q1/L	+15.72	+14.85	+14.82	−0.90	−0.03
E5/0	+11.23	+11.23	+10.37	−0.86	−0.86
E5/L	+20.85	+20.85	+20.85	+0.00	+0.00
Mean				−0.57 dB	−0.22 dB
Std. dev.				0.68 dB	0.48 dB

RR25E			
Set	Subjective	Calculated	Difference
D3/0	−5.45	−3.89	+1.56
D3/L	+0.15	+1.54	+1.39
E5/0	−4.98	−4.72	+0.26
E5/L	+1.58	+1.16	−0.42
F6/0	−1.55	−0.44	+1.11
F6/L	+7.00	+7.91	+0.91
G5/0	−6.67	−6.33	+0.34
G5/L	−0.23	+0.07	+0.30
Q1/0	−2.35	−2.18	+0.17
Q1/L	+6.48	+6.01	−0.47
R2/0	−8.80	−8.02	+0.78
R2/L	−1.13	−0.11	+1.02
T1/0	−2.98	−2.10	+0.88
T1/L	+1.88	+2.75	+0.87
U1/0	−5.25	−4.17	+1.08
U1/L	−1.03	−0.47	+0.56
V1/0	−8.95	−8.41	+0.54
V1/L	−1.75	−0.67	+1.08
Mean			+0.67 dB
Std. dev.			0.54 dB

SLR			
Set	Subjective	Calculated	Difference
D3/0	+4.45	+3.72	−0.73
D3/L	+5.75	+4.61	−1.14
D4/0	+2.55	+3.01	+0.46
D4/L	+3.65	+3.96	+0.31
F6/0	−0.55	+0.12	+0.67
F6/L	+8.95	+9.55	+0.60
F7/0	+1.10	+0.28	−0.82
F7/L	+10.20	+9.83	−0.37
T1/0	+7.80	+9.02	+1.22
T1/L	+12.00	+13.44	+1.44
T2/0	+10.45	+9.71	−0.74
T2/L	+15.00	+14.21	−0.79
U1/0	+5.55	+6.08	+0.53
U1/L	+10.30	+11.31	+1.01
Mean			+0.12 dB
Std. dev.			0.83 dB

RLR			
Set	Subjective	Calculated	Difference
D3/0	−6.85	−5.82	+1.03
D3/L	−1.45	−0.26	+1.19
D4/0	−9.15	−6.42	+2.73
D4/L	−3.75	−0.77	+2.98
F6/0	−2.45	−2.83	−0.38
F6/L	+5.55	+5.70	+0.15
F7/0	−1.50	−2.90	−1.40
F7/L	+6.05	+5.56	−0.49
T1/0	−5.55	−3.90	+1.65
T1/L	−0.30	+1.00	+1.30
T2/0	−3.75	−3.71	+0.04
T2/L	+2.25	+1.75	−0.50
U1/0	−6.85	−6.24	+0.61
U1/L	−2.30	−2.34	−0.04
Mean			+0.63 dB
Std. dev.			1.21 dB

TABLE 3-5 *(end)*

***d)* Comparison of subjective and calculated results using the ISO-532B algorithm**

SR25E					
Set	Subjective 1	Subjective 2	Calculated	Difference 1	Difference 2
F6/0	+3.35	+2.45	+0.33	−3.02	−2.12
F6/L	+10.97	+10.93	+10.62	−0.35	−0.31
Q1/0	+5.62	+5.33	+4.87	−0.75	−0.46
Q1/L	+15.72	+14.85	+14.79	−0.93	−0.06
E5/0	+11.23	+11.23	+9.74	−1.49	−1.49
E5/L	+20.85	+20.85	+20.49	−0.36	−0.36
Mean				−1.15 dB	−0.06 dB
Std. dev.				0.92 dB	1.05 dB

RR25E			
Set	RR25	Calculated	Difference
D3/0	−5.45	−4.33	+1.12
D3/L	+0.15	+1.14	+0.99
E5/0	−4.98	−5.14	−0.16
E5/L	+1.58	+0.70	−0.88
F6/0	−1.55	−1.61	−0.06
F6/L	+7.00	+6.95	−0.05
G5/0	−6.67	−6.78	−0.11
G5/L	−0.23	−0.35	−0.12
Q1/0	−2.35	−2.28	+0.07
Q1/L	+6.48	+5.78	−0.70
R2/0	−8.80	−8.35	+0.45
R2/L	−1.13	−0.54	+0.59
T1/0	−2.98	−2.39	+0.59
T1/L	+1.88	+2.45	+0.57
U1/0	−5.25	−4.45	+0.80
U1/L	−1.03	−0.75	+0.28
V1/0	−8.95	−8.83	+0.12
V1/L	−1.75	−1.16	+0.59
Mean			+0.23 dB
Std. dev.			0.52 dB

SLR			
Set	Subjective	Calculated	Difference
D3/0	+4.45	+4.13	−0.32
D3/L	+5.75	+5.03	−0.72
D4/0	+2.55	+3.36	+0.81
D4/L	+3.65	+4.34	+0.69
F6/0	−0.55	−0.86	−0.31
F6/L	+8.95	+9.21	+0.26
F7/0	+1.10	−0.70	−1.80
F7/L	+10.20	+9.56	−0.64
T1/0	+7.80	+9.15	+1.35
T1/L	+12.00	+13.72	+1.72
T2/0	+10.45	+9.89	−0.56
T2/L	+15.00	+14.53	−0.47
U1/0	+5.55	+5.79	+0.24
U1/L	+10.30	+11.48	+1.18
Mean			+0.10 dB
Std. dev.			0.93 dB

RLR			
Set	RLR	Calculated	Difference
D3/0	−6.85	−5.87	+0.98
D3/L	−1.45	+0.04	+1.49
D4/0	−9.15	−6.37	+2.78
D4/L	−3.75	−0.37	+3.38
F6/0	−2.45	−3.51	−1.06
F6/L	+5.55	+5.51	−0.04
F7/0	−1.50	−3.60	−2.10
F7/L	+6.05	+5.36	−0.69
T1/0	−5.55	−4.03	+1.52
T1/L	−0.30	+1.26	+1.56
T2/0	−3.75	−3.67	+0.08
T2/L	+2.25	+2.16	−0.09
U1/0	−6.85	−5.88	+0.97
U1/L	−2.30	−1.66	+0.64
Mean			+0.67 dB
Std. dev.			1.42 dB

TABLE 3-6

Summarized results showing the mean differences and standard deviations between the subjective and calculated R25Es and LRs using various algorithms

	SR25E		RR25E		SLR		RLR	
	Mean [a]	Std. dev. [a]	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Chinese	−0.57 −0.22	0.68 0.48	+0.67	0.54	+0.12	0.83	+0.63	1.21
P.XXE	−1.41 −1.06	0.64 0.45	−0.33	0.60	+0.12	0.82	+0.80	1.27
P.79 Cor	+1.56 +1.91	0.87 0.57	+1.83	0.79	−0.39	0.85	+0.01	1.11
ISO-532B	−1.15 −0.06	0.92 1.05	+0.23	0.52	+0.10	0.93	+0.67	1.42

[a] Two values in each block correspond to subjective test results quoted from different Technical Reports of the CCITT Laboratory.

Thus, the loudness loss becomes:

$$A = -\frac{10}{m} \log_{10} \int_{-\infty}^{\infty} K(X) \cdot 10^{-0.1mL(X)} \, \mathrm{d}X \tag{4-1}$$

where

m is the loudness growth factor

X $\log_{10} \{ F/F_0 \}$; $F_0 = 1$ kHz

$K(X)$ is the loudness weighting factors 4.2

$L(X)$ is the attenuation of the filter

For $K(X)$ it is stipulated that

$$\int_{-\infty}^{\infty} K(X) = 1 \tag{4-3}$$

Otherwise, $K(X)$ remains to be determined as well as the value of m. [In Equation (4-1) however, the exact value of m has only a second-order effect as has been discussed in other contributions.]

For a *high-pass* filter with negligible loss in the pass band and sharp cutoff at $F = F_c$ $(X = X_c)$ we get:

$$A_H = -\frac{10}{m} \log_{10} \int_{X_c}^{\infty} K(X) \, \mathrm{d}X \tag{4-4}$$

Similarly, for a *low-pass* filter

$$A_L = -\frac{10}{m} \log_{10} \int_{-\infty}^{X_c} K(X) \, \mathrm{d}X \tag{4-5}$$

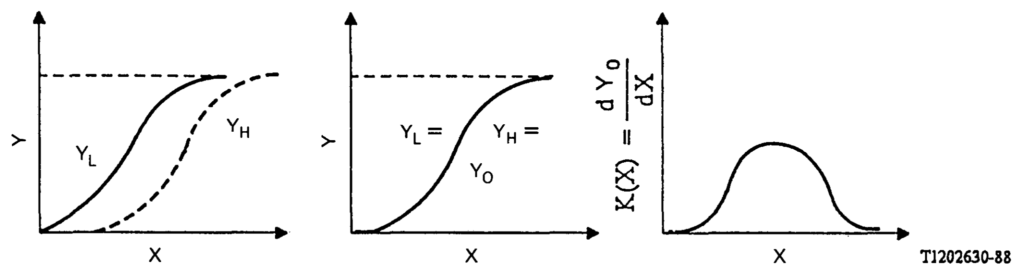
Using Equation (4-3) we get

$$Y = \int_{-\infty}^{X_c} K(X)\,\mathrm{d}X = 10^{-0.1mAL(X_c)} = 1 - 10^{-0.1mAH(X_c)} \tag{4-6}$$

For a chosen value of m, we may now plot as a function of X_c

$$\left.\begin{aligned} Y_L &= 10^{-0.1mAL(X_c)} \\ Y_H &= 1 - 10^{-0.1mAH(X_c)} \end{aligned}\right\} \tag{4-7}$$

S-shaped curves are obtained as in Figure 4-1 a). If the two curves more or less coincide as in Figure 4-1 b) the "best" value of m has been found. Then a mathematical expression for a curve Y_0 which fits the coincidence curve reasonably well is sought. The derivative of Y_0 thus gives $K(X)$.



a) Non-optimum value of m *b) Optimum value of m* *c) Derivation of K(X)*

FIGURE 4-1

Graphical derivation of loudness loss weighting factor K(X)

Of course S-shaped curves can be described by an infinite number of mathematical functions. However, the normal error integral turns out to be a suitable choice. Plotting $Y(A)$ on a "normal distribution diagram" paper gives, in essence, straight lines.

Figures 4-2, 4-3 and 4-4 present the results from data given in [4]. (The measurements were made at STL in January 1986, May 1975 and February 1975.) It is interesting to note how well the points cluster around straight lines, especially in Figure 4-2. The corresponding $K(X)$-curves are plotted in Figure 4-5 together with a curve derived from [5] as presented in [6].

4.3 *Discussion and conclusions*

It is remarkable that the weighting curves depicted in Figure 4-5 coincide so closely although they were made by very different test teams.

Curve 4 in Figure 4-5 has been used as a kind of reference in the further development of the "simplified" algorithm P.79A. The STL HP-LP measurements seem to confirm that this "weighting reference" is quite suitable. Thus, the P.79A algorithm will give a reasonable estimation of attenuation distortion and bandwidth limitation effects.

Another conclusion is that the loudness loss caused by attenuation distortion and bandwidth limitation can be explained by the simple loudness rating model without resorting to the Zwicker algorithm.

Curve 4 in Figure 4-5 was used to compute the corresponding 20-weights for the 1/3-octave frequencies in the series summation for the 0.1-8 kHz band, see [6]. These are shown in Figure 4-6 together with the equivalent K_i-values for P.79. As can be seen, the P.79 curve has some absurd peaks and gives more emphasis to lower frequencies and less to higher frequencies. Thus, P.79 can be expected to underestimate the effect of how attenuation slope as a function of frequency influences the loudness loss of a connection. This seems to be verified experimentally, as reported in [7].

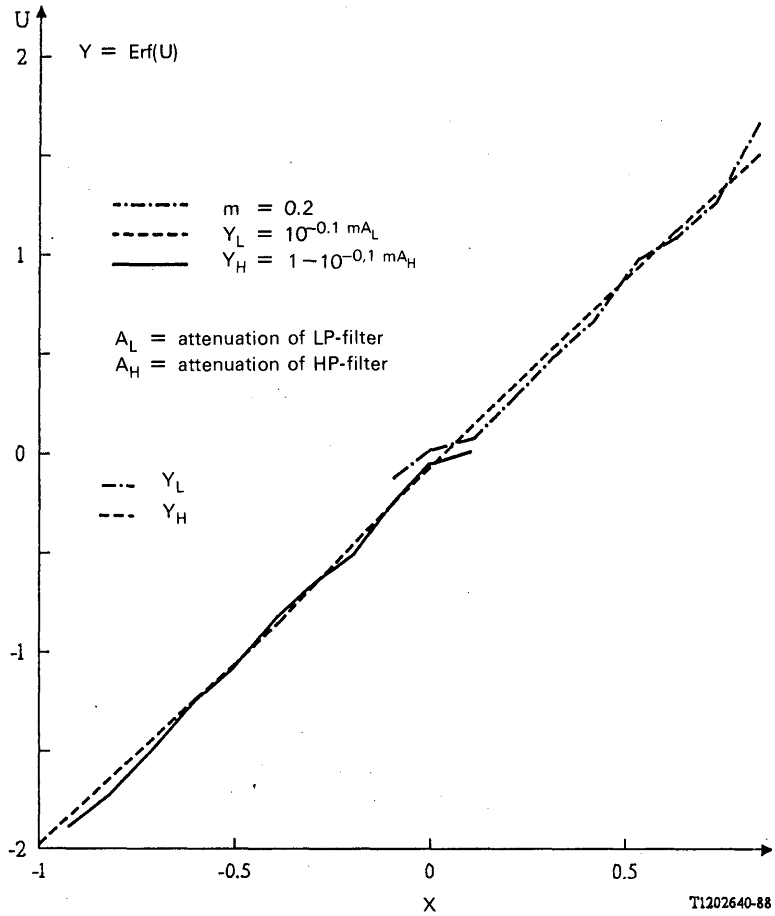


FIGURE 4-2

(From STL; January 1986)

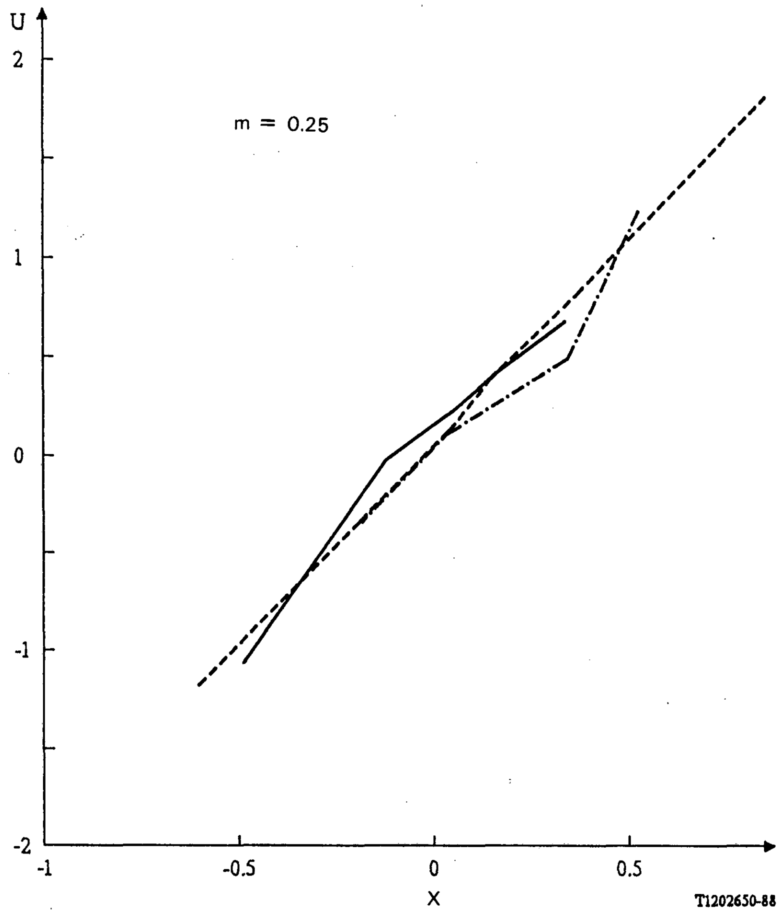


FIGURE 4-3
(From STL; May 1975)

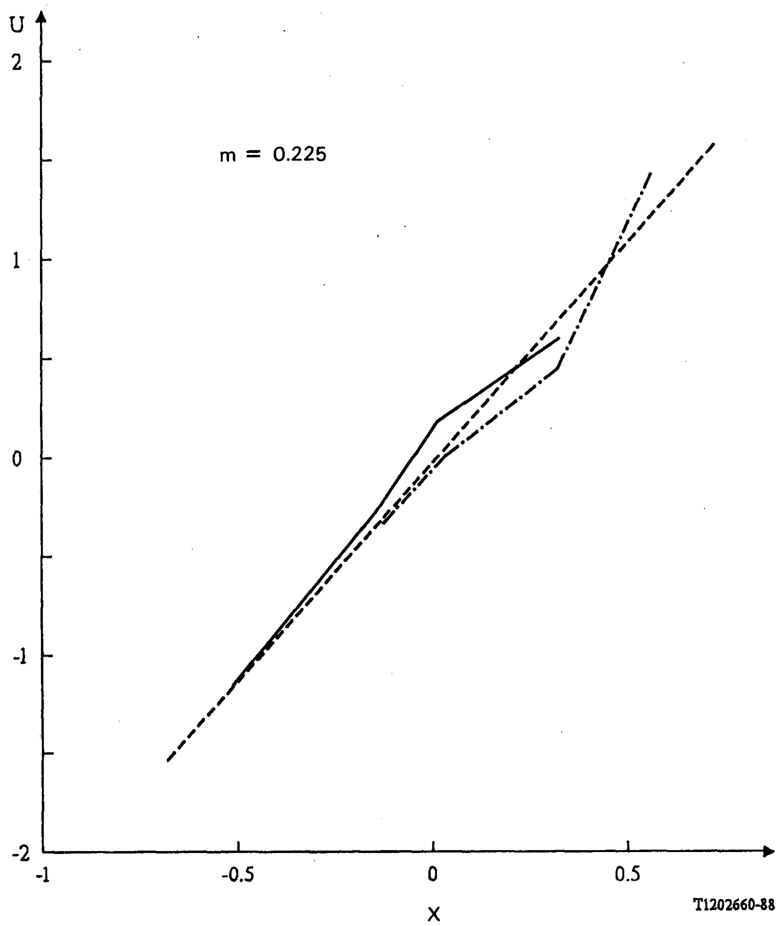


FIGURE 4-4

(From STL; February 1975)

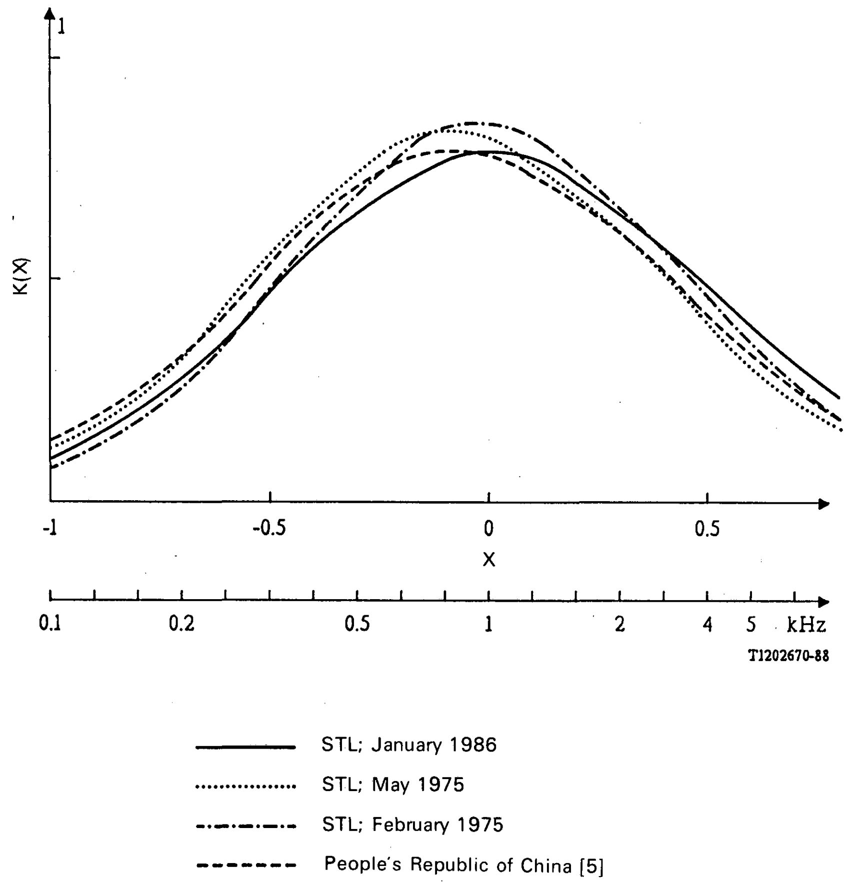


FIGURE 4-5

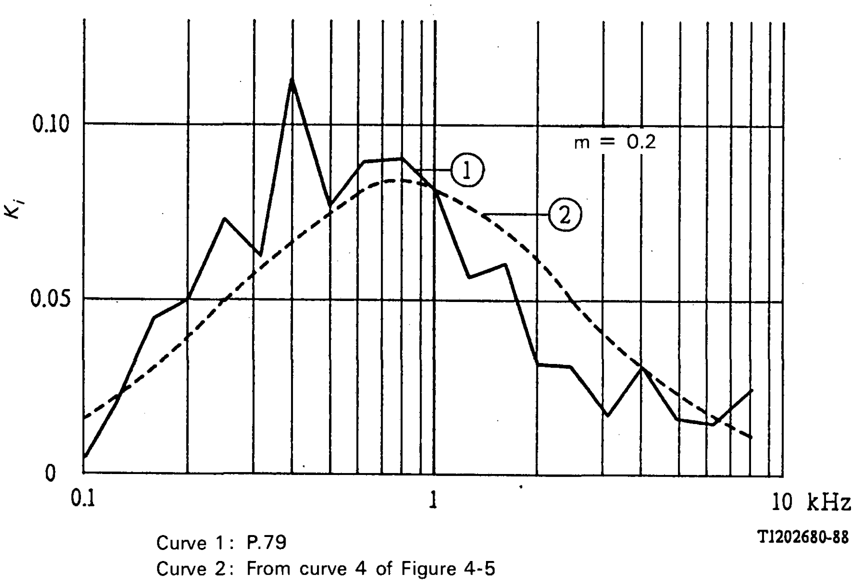


FIGURE 4-6

LR weighting coefficients K_i. 1/3-octave frequencies, 0.1-8 kHz

5 Loudness ratings and bandwidth in transmission planning (Contribution from ELLEMTEL, Sweden)

It is shown below that loudness ratings can be specified as "basic" parameters in the "common" band 0.3-3.4 kHz complemented with an *E*-factor for the band edges down to 0.2 kHz and up to 4 kHz. The *E*-factor can be determined numerically from attenuation values or by some simple network rules. The advantage of the method is a simplification for the transmission planner.

5.1 *Introduction*

Many Administrations seek to maintain good transmission properties in a telephone channel with a band of 200 to 4000 Hz, at least in the subscriber network. Under those circumstances it may seem natural to compute loudness ratings (LRs) using parameters specified for this band [8]. Because in this case the loss distortion is limited within the band, the additivity properties of the LRs will be satisfactory, i.e.:

$$OLR = SLR + JLR + RLR.$$

However, a connection may often contain links with an appreciable band edge attenuation distortion, virtually limiting the band to 300-3400 Hz. (This will be true for many interntional calls.) Such hard band-limiting corresponds to an increase of several decibels in LR. If LRs are computed for the band 0.1-8 kHz or even 0.2-4 kHz they can no longer be added without noticeable errors [9] which may cause confusion in the transmission planning.

In principle there are several ways of resolving the dilemma. The first is simply to ignore the improvement of a few dBs which some "wideband" local connections may possess. Thus, the American IEEE practice for objective loudness ratings is to use the band 0.3-3-4 kHz when computing (or measuring) the LRs.

The second method is to apply bandwidth correction factors to the LRs. One may compare with the CCITT concept of "corrected reference equivalents" which is tailored to the subscriber's actual loudness impression. A 200-4000 Hz circuit will have a lower CRE value than a 300-3400 circuit having the same midband loss. The effect of band-limiting is taken care of by correcting the wideband values by adding the so-called *D*-factors according to certain rules. [The CCITT *D*-factors may not be quite correct, however, as they were derived from measurements using SRAEN filters. These are not truly representative of modern transmission circuits [10].]

Considering modern trends of trying to improve the telephone channel's low-frequency response, it seems appropriate for the LR calculations to use a "wideband" method with corrections.

Such a methodology will be described below and this can be applied to transmission planning.

The LRs are basically calculated for the narrow "common band" 0.3-3.4 kHz. These LRs can be added without loss of accuracy. A correction, the *E*-factor, is computed for the transmission at the band edges. The *E*-factor is subtracted from the "common band" OLR to obtain the "wideband" $OLR\ (W)$.

5.2 *The E-factor as a band edge correction of LR*

In general, a loudness rating can be thought of as a "frequency-weighted average" of an electro-acoustical attenuation.

According to Recommendation P.79 the electro-acoustical properties should be evaluated in the band 0.1-8 kHz. For practical reasons the computations are often limited to the band 0.2-4 kHz. (The W_i-weights are then diminished by 0.3 dB). However, only in the band 0.3-3.4 is one assured of a real transfer of signals under all circumstances. At the band edges, 0.2 to 0.3 and 3.4 to 4 kHz, the attenuation of a specific link in a connection may be so high as virtually to stop transmission. This could result in a reduction of several decibels in a subjectively measured loudness impression of a voice signal.

To handle this properly it is convenient to characterize the electro-acoustical attenuations separately for the "common band" 0.3-3.4 kHz and for the band edges.

In the common band each link is characterized by the weighted average of the electroacoustical loss, i.e. SLR, RLR or JLR, and the LRs can be added. For example, for the circuit as shown in Figure 5-1, consisting of two telephone sets interconnected via a number of transmssion links, the following relation should hold at any interfacte P between the links:

$$\begin{aligned} SLR_P &= SLR_0 + \sum_{i=1}^{P} JLR_i \\ RLR_P &= \sum_{i=P+1}^{P} JLR_i + RLR_0 \end{aligned} \qquad (5\text{-}1)$$

(Any mismatch attenuation effects at the interfaces can be treated as special forms of JLRs).

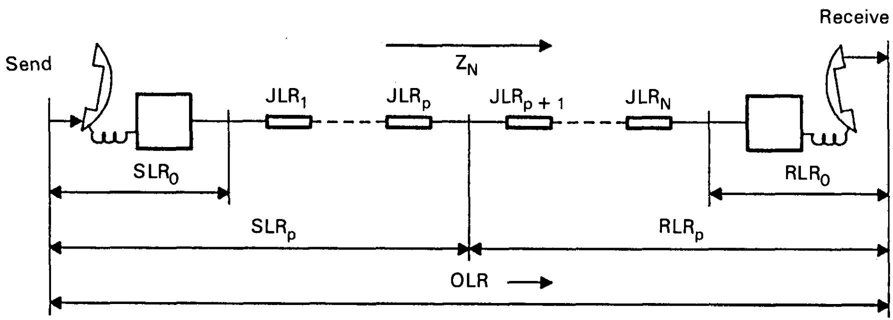


FIGURE 5-1

LRs for links in tandem

At the band edges the connection is characterized by its ability to transmit voice signals, i.e. the *E*-factor. Zero band edge losses means $E = 2.5$ dB. (Details will be given later).

For a complete connection as shown in Figure 5-1, the overall loudness rating is:

In the common band 0.3-3.4 kHz $OLR = SLR + RLR$ (5.2)

In the full band 0.2-4 kHz $OLR(W) = OLR - E$ (5-3)

In the following, the general mathematical expressions for the LRs and the *E*-factor are given. It is shown how to apply them to telephone sets and various transmission links.

The *E*-factor may be designated the "loudness improvement".

5.3 *General mathematical expressions*

In the common band 0.3-3.4 kHz the general LR algorithm can be written as:

$$LR = L_0 + \overline{L} \qquad (5\text{-}4)$$

$$\overline{L} = -\frac{10}{m} \log_{10} \sum_{i=1}^{N} K_i \cdot 10^{-0.1 m L_i} \qquad (5\text{-}5)$$

(the summation being made for $f_i = 0.315 \dots 3.15$, the 1/3-octave ISO frequencies)

when

L_i	are the values of electroacoustic loss for the LR in question
L_0, K_i, m	are constants to be specified below.

Note – Equations 5-4 and 5-5 are mathematically equivalent to the "W_i-algorithm" as explained in [11] but are more convenient to use in the following.

When the spread between minimum and maximum values of the L_i's is moderate, the following expression can be used for $\overline{L}$.

$$\overline{L} = \sum_{i=1}^{11} K_i L_i \tag{5-6}$$

The full band 0.2-4 kHz overall loudness rating was given by Equation 5-3, i.e.:

$$OLR(W) = OLR - E$$

The expression for the E-factor is:

$$E = C_1 \cdot 10^{-0.1m(L_{01}-\overline{L})} + C_2 \cdot 10^{-0.1m(L_{02}-\overline{L})} + C_3 \cdot 10^{-0.1m(L_{03}-\overline{L})} \tag{5-7}$$

where

L_{01}, L_{02}, L_{03} are the band edge losses at 0,2, 0,25 et 4 kHz respectively.

C_1, C_2 and C_3 are constants to be specified below (for the derivation, see Annex A).

The constants of Equations 5-4 and 5-5 may in principle be derived from any defined LR algorithms. However, Recommendation P.79 is less suitable for use in the context of transmission planning because of a lack of accuracy as discussed in [11] and [7]. The very simple algorithm, designated by "C" in [11], seems to be just as accurate as any other investigated so far and is therefore chosen here. (It also has the advantage of closely resembling the IEEE objective loudness rating.)

The constants used in Equations (5-4), (5-5), (5-6) and (5-7) are given in Tables 5-1 and 5-2.

TABLE 5-1

Constants used in equations (5-4), (5-5) and (5-6)

$K_i = 0.05$ for $f_i =$ 0.315 and 3.15 kHz
$K_i = 0.1$ for $f_i =$ 0.4, 0.5 ... 2, 2.5 kHz
$m = 0.2$

LR	SLR	RLR	OLR	JLR
L_0	−3	12	9	0

TABLE 5-2

Constants used in equation (5-7)

$C_1 = 0.5$	$C_2 = 1$	$C_3 = 1$

A perfectly flat mouth-to-ear acoustic frequency response in the band 0.2-4 kHz will thus result in E = 0.5 + 1 + 1 = 2.5 dB, i.e. *OLR*(*W*) is 2.5 dB lower than *OLR* for a "flat" channel limited to 0.3-3.4 kHz.

In the following the *E*-factor is computed for a number of cases including "broadband" and "narrowband" telephone sets in combination with different types of transmission links. It turns out that some rather simple rules can be set up for the approximate determination of the *E*-factor.

5.4 *Telephone sets*

Suppose the transmission channel between the sending and receiving telephone sets is flat within the band 0.2-4 kHz. Then the *E*-factor, the loudness improvement, characterizes the bandwidth performance of the sets. Let this be designated E_T. Table 5-3 gives some examples. It is worth noticing that the spread in *E* around the average value 1.3 is quite moderate.

TABLE 5-3

Examples of E-factors for some telephone sets

Type of set	E_T
1) Old-type carbon microphone	1.9
2) Old-type carbon microphone	1.5
3) Old-type carbon microphone	1.1
4) W.E. type 500	1.3
5) Electret microphone	1.8
6) Digital set; new specification	0.8
7) Average of 90 types of sets	1.3

5.5 *Transmission links*

To characterize the loudness improvement performance of the transmission links as such, it is convenient to compute the *E*-factor under the assumption that the telephone sets have a flat frequency response in the band 0.2-4 kHz. Let this transmission channel *E*-factor be designated E_C. In § 5.6 the resulting *E*-factor will be given for various typical combinations of E_T and E_C.

When connecting several transmission links in tandem, mismatch may occur. These effects can be diminished, however, by using complex nominal impedances in the subscriber networks, as many Administrations already do.

In general, mismatch losses can be considered by computing their *JLRs*.

The "common band" performance of the links are characterized by $JLR = \overline{L}$ according to Equations (5-4), (5-5) and Table 5-1. As large attenuation distortions within this band are not allowed, the very simple Equation (5-6) can be used for computing $\overline{L}$. (It is interesting to note that this corresponds in effect to averaging the loss over a log (*f*)-scale, a method which has been verified empirically a long time ago).

When several links are connected in tandem, E_C can of course be computed from the total band edge losses. However, some simple approximate rules can be used for the combination of individual E_C-factors for *JRL*.

5.5.1 *Subscriber cables*

Surprisingly, the typical loss curve of a non-loaded subscriber cable produces the same loudness improvement as a full-bandwidth channel, i.e. $E_C = 2.5$ dB. This is due to the fact that at the lower band edge the loss is lower than the average $\overline{L}$ which compensates for the higher loss at the upper band edge.

When a subscriber cable is connected in tandem with a narrow band device, it turns out that the E_C-factor for that device applies.

5.5.2 *Band-limiting equipment*

Band-limiting in a telephone connection can be caused by heavily loaded subscriber cables, FDM and PCM equipment. Figure 5-2 shows some idealized attenuation curves for which the E_C-factors have been computed.

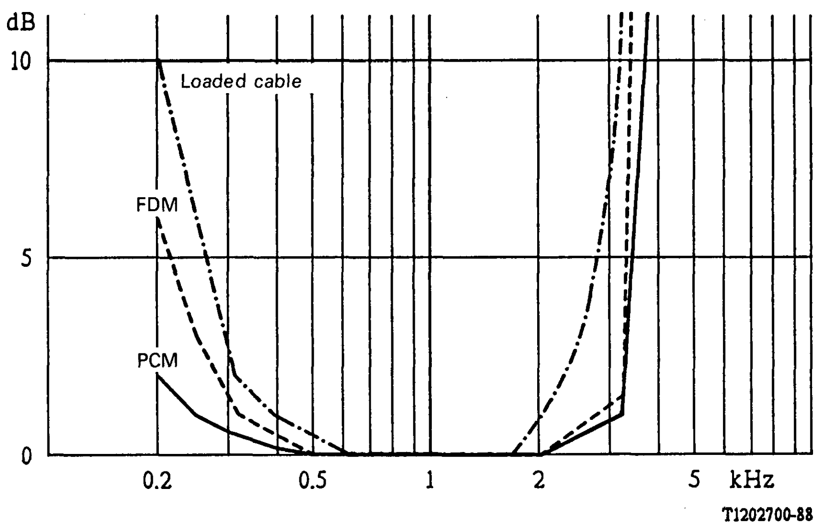


FIGURE 5-2

Examples of band-limiting equipment: idealized loss curves for PCM, FDM and a heavily loaded subscriber cable

For the attenuation curves in Figure 5-2 the following E_C-values are obtained:

	E_C
Heavily loaded subscriber cable	1.2 dB
1 FDM link	1.4 dB
1 PCM link	1.9 dB

When several PCM and FDM links are connected in tandem the E_C-values according to Table 5-4 are obtained.

5.6 *Complete connections*

The loudness improvement, the *E*-factor, has been computed for a number of combinations of telephone sets and transmission links. For each telephone characteristic the "total" *E*-factor has been plotted against the "channel" E_C-factor. The results are presented in Figures 5-3 and 5-4.

TABLE 5-4

E_C for PCM and FDM links in tandem

No. of FDM links \ No. of PCM links	0	1	2	3	4	5
0	2.5	1.9	1.6	1.4	1.3	1.2
1	1.4	1.3	1.2	1.1	1.0	1.0
2	1.1	1.0	1.0	0.9	0.9	0.8
3	0.9	0.9	0.8	0.8	0.7	0.7
4	0.8	0.7	0.7	0.7	0.6	0.6
5	0.7	0.6	0.6	0.6	0.5	0.5

Figure 5-3 shows the *E*-factor for the "average" analog and the "digital" telephone set. (The "average" was taken as the mean of 90 different types of commercial sets. The "digital" corresponds to the new CCITT specification for digital sets).

Figure 5-4 illustrates the spread in the *E*-factor for a number of widely varying analog telephone characteristics. (It is worth noticing that the spread, after all, is fairly moderate).

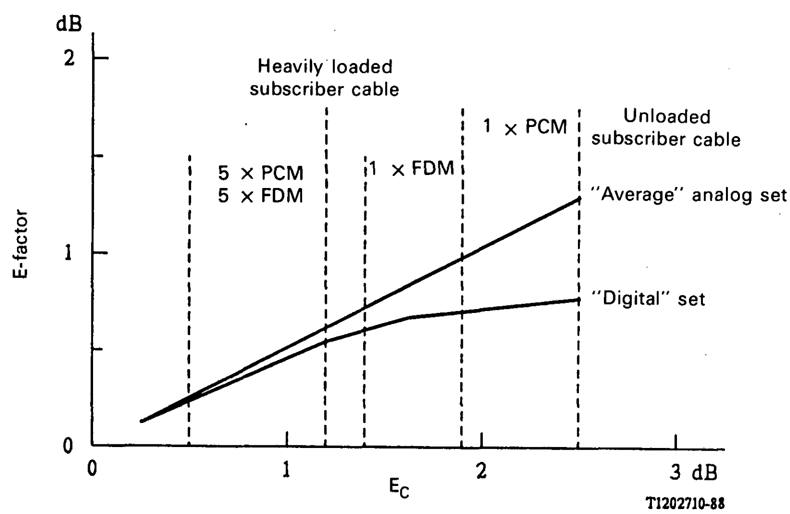


FIGURE 5-3

The E-factor as a function of channel E_C for "average" analog and digital sets

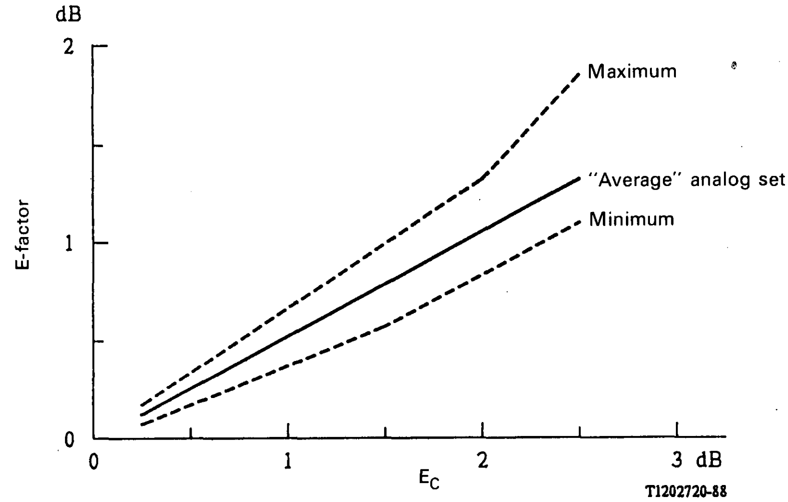


FIGURE 5-4

The E-factor as a function of channel E_C for "typical" analog sets

Considering the general requirements of transmission planning, there is hardly a need to specify the loudness improvement, the E-factor, more accurately than within steps of 0.5 dB. Therefore, instead of calculating the E-value in each application, one can follow the rules given in Tables 5-5 and 5-6 for analog and digital sets respectively.

TABLE 5-5

E-factor, analog sets

E	Links in tandem
1.5	Subscriber cable, non-loaded
1.0	1 × PCM, . . ., 3 × PCM 1 × FDM
0.5	Subscriber cable, heavy coding
	4 × PCM 2 × FDM
0	5 × PCM + 5 × FDM

Note – Non-loaded subscriber cable sections do not affect the *E*-factor.

TABLE 5-6

E-factor, digital sets

E	Links in tandem
1.0	All digital connection
0.5	1 D/A-A/D to 6 D/A-A/D connections
0	7 D/A-A/D connections

5.7 *Conclusions*

The transmission planner can obtain loudness ratings by quite simple numerical methods: computing and adding individual LRs for the "common band" 0.3-3.4 kHz and correcting for the band edge transmission by subtracting the *E*-factor. The *E*-factor can be determined by some uncomplicated rules.

The results can be expected to be more accurate than calculations based on Recommendation P.79.

6 **Algorithms for transmission planning** (from Ellemtel, Sweden)

This Section compares loudness ratings calculated by algorithms as defined in the IEEE, the OREM-8 and the Recommendation P.79 standards with P.79A, an adaptation of Recommendation P.79 for transmission planning. The computations encompassed 71 different telephone set characteristics and 27 cases of subscriber cable attenuation distortions.

The results show that, in typical transmission planning situations, the loudness ratings by P.79A are closely related to the others. In addition, P79A has the best additivity accuracy and the simplest form. Administrations are therefore urged to make some comparative studies using this P.79A algorithm in transmission planning.

6.1 *Introduction*

Some preliminary comparisons between different loudness loss estimation methods for telephone network transmission planning are presented (sidetone algorithms are not treated, however). The methods discussed are all based on objective measurements on telephone instruments.

The transmission planner has some basic requirements:

1) to obtain a meaningful and reliable indication of the telephone speech transmission quality as regards the acoustic loss between the talking subscriber's mouth and the listening subscriber's ear;

2) to characterize the individual parts of a connection by numerical values in such a way that their sum equals the acoustic loss measure of the complete connection;

3) to ensure that the numerical values can be determined reliably, inexpensively and with sufficient accuracy from measurements.

The third requirement means that for quality control and day-to-day use only *objective* electro-acoustic measurements can be considered. Of course, subjective measurements have been absolutely necessary in the past to establish the general level of reference. But they are more costly to carry out and give far less repeatable results even under favourable conditions. For example the CCITT Laboratory has periodically made subjective reference equivalence measurements on the same stable telephone set since late 1981. The values seem to wander slowly with time within a 3 dB range. See [12].

6.2 *Different loudness rating methods*

Objective electroacoustic loss (or sensitivity) measurements form the basis of several planning methods used by Administrations for their national networks. The most important of these "loudness ratings" are listed below, in the order most widely applied.

1) *IEEE standard.* This is now used in the USA and Canada. It is well-documented and reliably instrumented (see § 1 and [1]).

2) *OREM-B standard* (in different versions). The version worth comparing to is currently being more accurately redefined by a standardization body in the Federal Republic of Germany.

3) *Reference equivalents,* but determined objectively by some kind of instrumentation. It should be noted that the "reference loss" of a junction circuit is often determined as the dB average over a log(f)-scale.

4) Recommendation P.79. This is the "youngest" standard. The U.K. has the widest experience in its application. From a transmission planner's point of view it leaves something to be desired, and a possible modification is described in [13].

Of course, to ensure telephone speech loudness quality in international connections it is desirable that Administrations employ methods for loudness rating evaluations which are compatible, or preferably identical. The aim of this exposition is to show that for transmission planning a modified, or amended, version of Recommendation P.79 is compatible with the IEEE, the OREM-B and the original Rec. P.79 standard, as well as with some aspects of the reference equivalent practice. In addition, this amended P.79 algorithm needs much less computation effort. (The principles are given in [13] and in § 5).

For a transmission planner the most important loudness parameters are the send, the receive and the junction loudness ratings. Note that the send and receive loudness ratings for transmission planning purposes *do not* have to relate uniquely to subjective values determined by comparison with some reference telephone system set-up. The important aspect is that the (send + receive + junction) loudness ratings add up to the correct overall loudness rating (OLR), which in turn should have a good correlation with a subjective measure. (But this cannot mean that different countries can use different ratings that subdivide the OLR differently!)

What order of computation accuracy is needed? With regard to the additivity of loudness ratings as well as for setting limits in a national transmission plan, 0.5 dB accuracy would probably be adequate. For translating between different standards, 1.5 dB seems sufficient. (The loudness rating manufacturing tolerances for telephone sets, however, have to be much wider).

6.3 *Basics of the loudness rating methods*

The physiological background to the loudness rating methods is as follows:

The ear functions as a bank of narrow bandpass filters. (About 64 equivalent filters having bandwidths of 50-500 Hz, called "critical frequency bands"). In a specific band only, that portion of the sound energy which lies above the hearing threshold is presented to the brain as a stimulus. The brain combines all filter outputs of a complex sound (such as the human voice) into an impression of loudness.

Each filter contributes in proportion to an exponent m of its output power, m being in the order of 0.2-0.3 for normal speech levels. For very low levels, near the threshold of hearing, as well as for very high levels, m equals 1, corresponding to power addition. But it is important to note that for normal telephone speech, weaker spectral components contribute more to the total loudness impression than if they were added on a direct power basis. As a matter of fact, they approximately add on a linear dB scale.

A more complete theory of hearing, according to Zwicker, also takes into account certain masking effects between the stimuli from the different frequency bands. For telephone purposes, however, the Zwicker method does not seem to give any better agreement with subjective values. See [11] and [7].

In an active telephone connection the listener's impression of loudness is influenced by his own hearing sensitivity/frequency curve and the talker's speech spectrum. A general loudness rating thus has to consider the statistical average for the human population.

The loudness rating algorithms mentioned can all be written in the following form:

$$LR = L_0 + \overline{L} \tag{6-1a}$$

$$\overline{L} = -\frac{10}{m} \log_{10} \sum_{i=1}^{N} K_i \cdot 10^{-0.1mL_i} \tag{6-1b}$$

where

$$\sum_{i=1}^{N} K_i = 1 \tag{6-2}$$

L_0, m and K_i are constants. The index "i" refers to the frequency f_i.

L_i is the appropriate measured and/or computed electroacoustic loss in dB for send, receive, overall and junction rating respectively.

The coefficients K_i depend on the "average" speech and hearing properties, some average transmission characteristics of a typical circuit, as well as the spacing between the frequencies f_i. Thus the K_i's are in general frequency-dependent with proportionally lower values at the band edges. (An example of how to determine the K_i's is found in [6]). However, the exact form of the K_i variation is not very critical, as will be shown in the following.

A useful insight into the mathematical behaviour of the loudness rating formula can be had from a Taylor series expansion. Put

$$L_m = \sum_{i=1}^{N} K_i \cdot L_i \tag{6-3}$$

Equation (6-1b) can now be written as

$$\overline{L} = L_m - a \sum_{i=1}^{N} K_i (L_i - L_m)^2 \tag{6-4}$$

ignoring higher-order terms. Here

$$a = \frac{m}{20} \ln (10) = 0.115\ m \tag{6-5}$$

The coefficient a is small, in the range 0.02 to 0.03 for m in the range 0.175 to 0.3.

Thus, the exact value of m is quite uncritical, as will be shown numerically below. As a matter of fact, the second term in equation (6-4) can be almost disregarded if the L_i's deviate only moderately from the average L_m, which is the case for most telephone send and receive characteristics.

This explains why send, receive and junction loudness ratings can be added to get the overall loudness rating, at least as long as no hard band-limiting is included in the frequency range.

Moderate variations in the coefficients K_i between diferent algorithms can of course be interpreted as small correction terms to be added to the L_i's. Thus, the results from different algorithms can be expected to differ only by some constants, fairly independent of the telephone set characteristics.

Similarly, other corrections involving small frequency-dependence can also be treated as constants added to the loudness ratings (for example earcap leakage, difference between measuring set-ups, etc.).

In practice, the transmission bandwidth of a telephone connection may of course differ, depending on what type of links are included. Discussions of how this can be handled when determining the loudness rating can be found in [13] and in § 5, which propose an amendment to Rec. P.79.

To investigate the properties of various loudness rating algorithms it is useful to make computations for a number of typical telephone set characteristics. Below, seventy-one widely different characteristics have been used to obtain a statistical picture. See Annex B for details.

6.4 *Particulars about some algorithms*

Only the mathematics of the algorithms are dealt with. Differences in the measuring arrangements are not considered here. The statistics refer to computations using the 71 different set characteristics as mentioned above. The result of the computations are given with 0.01 dB precision only to show the differences clearly. This "accuracy" is of course not at all necessary or useful in practice.

6.4.1 *IEEE standard*

The computation range is 0.3 to 3.4 kHz with $m = 0.22$. The K_i-weighting is flat on a log (f)-scale, i.e. all K_i's are equal except the end points which have half the value. (From a physiological point of view the weighting should really taper off slightly at the band edges).

In the calculations presented here, 43 frequency points have been used.

6.4.2 *OREM-B standard*

Measurements of OREM values have traditionally been made with special electroacoustic instruments. Recently, however, FTZ in the Federal Republic of Germany has found it possible to compute OREM-B values according to Equation (6-1) using measured sensitivity curves [14].

The computation range is 0.2 to 4 kHz with $m = 0.3$. The K_i-weighting is flat on a log (f)-scale as for the IEEE method, but for the send and overall loudness ratings the loss of a SFERT filter is added to L_i in Equation (6-1). Fifty-three frequency points are used.

6.4.3 *Rec. P.79 standard*

The computation range most often used is 0.2 to 4 kHz with m = 0.175. The constants L_0 and K_i are transformed into the so-called W_i-weights which are added to the L_i's. The earcap leakage L_E is included for RLR and OLR.

The W_i's are of course tabulated in Recommendation P.79 and the corresponding K_i's are shown in diagrams in [7]. Fourteen frequency points, spaced ⅓ octave apart, are used.

The Rec. P.79 constants were tailored to make computed loudness ratings agree with the correspondings subjective values obtained by a *specific* CCITT test team some years ago. Therefore, in some respects Rec. P.79 does not quite represent subjective values obtained by "ordinary" people. Rec. P.79 puts too much emphasis on the lower frequencies and the weighting coefficients show some peculiar irregularities. (See [11] and [7] for a discussion.) Also, if a connection contains links of different bandwidths some ambiguity may occur.

When applying Rec. P.79 in practice, computer programs are most often used. Telephone set sensitivity curves and chain matrix data of the individual links are the inputs, and SLR and RLR are computed to an interface terminated by 600 ohms (which, incidentally, may not be the nominal impedance at that point!). JLR is not calculated very often.

6.4.4 *An amended Rec. P.79 standard: P.79A*

Details of this proposal are given in [13].

Eleven frequency points spaced ⅓ octave apart, are used. The computation frequencies start at 0.315 and end at 3.15 kHz, making the range extend virtually from 0.3 to 3.4 kHz.

The K_i's are all equal to 0.1 except for the end points where they are equal to 0.05. This trapezoidal weighting thus takes some account of the physiological facts about speech spectrum and hearing sensitivity curves.

Loudness rating for tandem-connected links of wider bandwidths than 0.3-3.4 kHz is taken care of by the "loudness-improvement" E-factor, which is subtracted from OLR.

In the more "exact" version of the amendment, $m = 0.2$. This is here designated "P.79A1".

A simpler approach is to let $m = 0$, i.e. use the weighted loss average according to Equation (6-3). This version is here designated "P.79A".

In Equation (6-1) the constant L_0 is:

for SLR: -3

for RLR: 12

for OLR: 9

for JLR: 0

When using P.79A1 or P.79A there is much less need for complex computer programs. The LRs of the individual links can be added with good accuracy.

6.4.5 *Dependence on the m-value*

The designation "d" is used to illustrate the change of a loudness rating as a function of the coefficient m. The "norm" value of a particular standard corresponds to $d = 0$.

It can be seen from Table 6-1 that the choice of m is very uncritical. It explains why P.79A1 with $m = 0.2$ and P.79A with $m = 0$ are fairly equivalent.

TABLE 6-1

Values of d (i.e. changes in LR) as a function of m

(Statistics from 71 telephone set characteristics)

Standard	m	Send		Receive	
		Mean	Standard deviation	Mean	Standard deviation
IEEE	0.175 0.2 0.22	−0.15 −0.05 0	0.05 0.02 0	−0.22 −0.08 0	0.02 0.01 0
OREM-B	0.175 0.2 0.3	−0.21 −0.13 0	0.13 0.10 0	−0.22 −0.14 0	0.11 0.08 0
P.79A1	0.02 0.175 0.2	0.33 0.05 0	0.21 0.03 0	0.07 0.01 0	0.07 0.01 0

6.4.6 *Additivity properties*

An indication of how well a certain algorithm is suited for planning purposes is how closely the sum of send, junction and receive ratings corresponds to the overall rating. The difference between OLR and the sum of SLR + JLR + RLR is denoted as "*D*".

Of special interest is how links of different bandwidths are treated. In a local, analog network the transmission range can be considered as 0.2 to 4 kHz. In a long-distance network (including international links) one can hardly be assured of transmission outside the range 0.3-3.4 kHz. Modern PCM systems, digital exchanges, etc., may have an effective band of 0.2-3.4 kHz.

The IEEE standard only deals with the "narrow band" 0.3-3.4 kHz.

The amended P.79 standard (P.79A1 and P.79A) uses the same "narrow band" for SLR, RLR and JLR calculations. If the actual transmission band is wider, the sum is diminished by a "loudness improvement factor" *E* to obtain an OLR which has a good correlation with subjective loudness impressions. (See [13] and § 5.)

The unmodified P.79 standard for practical cases uses the range 0.2-4 kHz. Depending on the bandwidths of the links and the choice of interface to which the SLR and RLR calculations refer, some ambiguity can then occur.

In Table 6-2 values of *D* are given and, for reasons of simplicity, the junction is considered to have a flat frequency response over the frequency band considered.

Note the large errors when P.79 is applied to a "narrow band" connection such as can be expected in international calls.

The IEEE standard has the second worst additivity performance. (But it is of course satisfactory in practice.)

TABLE 6-2

Values of D = OLR-SLR-RLR; JLR = 0

(Statistics from 71 telephone set characteristics)

Algorithm	Band (kHz)	D (mean)	Standard deviation	D (max.)	D (min.)
IEEE	0.3-3.4	0.84	0.24	1.42	0.08
OREM-B	0.2-4	0.19	0.24	0.62	−0.48
P.79	0.2-4	−0.48	0.19	−0.11	−0.98
P.79	0.2-3.4	−0.12	0.13	0.17	0.56
P.79	0.3-3.4	−1.78	0.09	−1.60	−2.09
P.79A1	0.3-3.4	−0.06	0.20	0.45	−0.70
P.79A	0.3-3.4	0	0	0	0

6.5 *Numerical comparisons between loudness ratings of different standards*

6.5.1 *P.79 algorithm*

From a transmission planner's point of view, the use of the simple loudness rating algorithm P.79A seems attractive. It is then of special interest to compare with results obtained when applying the "normal" P.79 algorithm.

Table 6-3 shows the differences for SLR, RLR and OLR for connections having different bandwidths. (The frequency response is assumed to be flat within the passband, however.)

TABLE 6-3

Values of D = LR(P.79)-LR(P.79A)

(Statistics from 71 telephone set characteristics)

LR	Band (kHz)	D (mean)	Standard deviation	D (max.)	D (min.)
SLR	0.2-4 0.2-3.4 0.3-3.4	−1.12 −0.93 1.0	0.53 0.55 0.38	0.15 0.32 1.68	−2.61 −2.47 −0.03
RLR	0.2-4 0.2-3.4 0.3-3.4	−0.65 −0.46 1.52	0.46 0.47 0.31	0.43 0.61 2.06	1.95 −1.73 0.55
OLR	0.2-4 0.2-3.4 0.3-3.4	−0.85 −0.77 0.75	0.69 0.71 0.54	0.58 0.72 1.78	−2.57 −2.54 −0.59

It can be seen from Table 6-3 that the P.79 and the P.79A algorithms are reasonably equal to each other, considering the bandwidth ambiguity of the P.79. They seem to give the same numerical values (on the average) for a connection having a slightly narrower effective band than 0.2-3.4 khz but broader than 0.3-3.4 kHz, i.e. a case often to be expected in practice.

As mentioned earlier, the junction loudness rating, JLR is of less importance when using P.79 in practice. As a matter of fact JLR (P.79) can give somewhat misleading results because of the band edge peculiarities of the algorithm as discussed earlier.

JLR values calculated by the P.79A1 and P.79A algorithms are more useful to the network planner for several reasons. The additivity properties are better, and some previous investigations indicate good agreement with subjective measurements of loudness loss ([15]). (As a matter of fact the Swedish Administration has used similar algorithms for 20 years.)

The more "complete" algorithm P.79A1 may be safely assumed to give the closest agreement with subjective measurements of JLR. For the transmission planner it is of interest to compare with:

a) JLR results when using the still simpler algorithm P.79A.

b) Circuit losses as defined by the difference in relative levels ($= L_1$).

When the frequency response curve is flat all these quantities are of course identical. But typical unloaded subscriber cable introduces a high degree of attenuation distortion in the telephone channel. A number of cases (27) have been investigated, including different cable diameters, d.c. resistance (lengths) and terminations.

For these investigations the cable data were: capacitance 45 nF/km; diameters 0.4, 0.5 and 0.7 mm; lengths corresponding to 300, 600 and 1200 ohms d.c. resistance. The terminations were 600, 900 ohms and a typical complex impedance (200 ohms in series with a parallel combination 820 ohms and 115 nF).

The difference of JLR calculated using two algorithms is presented in Figure 6-1 as a function of the cable attenuation distortion, i.e. the difference between the loss at 4 kHz and the loss at 0.2 kHz. (For the same length of a cable, the complex impedance termination gives the largest distortion. In practice, distortions larger than about 15 dB should be avoided for various reasons.)

As can be seen from Figure 6-1 the P.79A algorithm is sufficiently accurate to be used for JLR calculations, the differences being less than 0.3 dB for a distortion up to 15 dB.

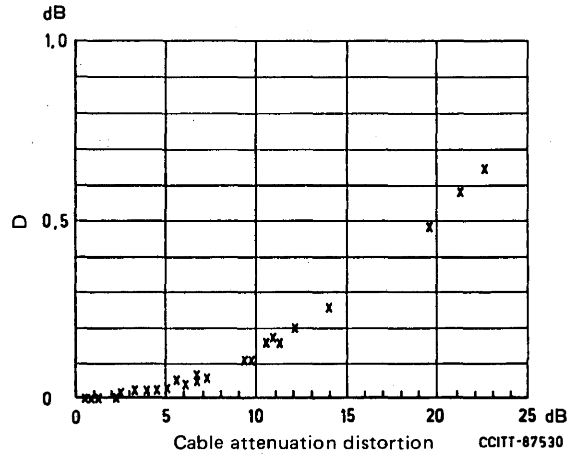


FIGURE 6-1

D = JLR(P.79A) – JLR(P.79A1) for subscriber cables

The subscriber cable loss as defined by L_1, the difference in relative levels, is simply the (composite) loss at 1 kHz according to the CCITT definition (even for a complex impedance termination). Figure 6-2 shows that the difference to the "true" JLR value is always less than 1 dB, and presumably less than 0.5 dB in the majority of practical cases.

Thus the difference in relative levels, L_1, can also be used, with good accuracy, as a measure of the change in loudness rating, JLR. This makes the task easier for the network planner.

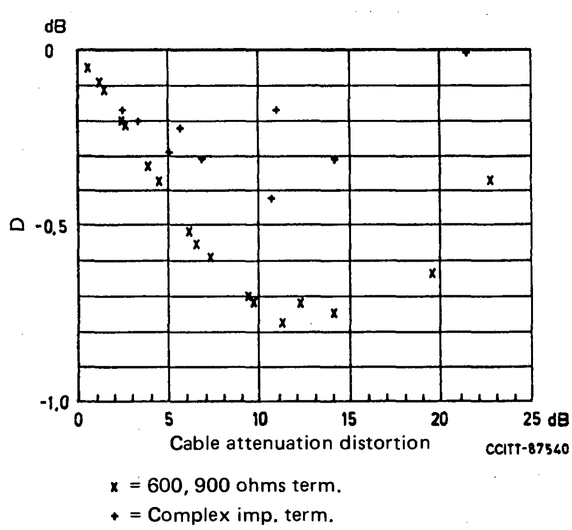


FIGURE 6-2

D = L_1 –JLR(P.79A1) for subscriber cables.
L_1, the difference in relative levels, equals the loss at 1 kHz

6.5.2 *IEEE and OREM-B algorithms*

To what degree will it be possible to make simple conversions between P.79A and the IEEE and OREM-B standards? That is, what is the difference

$$D = LR - LR(\text{P.79A}) \ ? \tag{6-6}$$

Of interest are the average D(mean) and the standard deviation for typical telephone set characteristics. Because possible differences in measuring set-ups have not yet been investigated, it is for the moment only relevant to study the standard deviations.

The computation results are presented in Table 6-4. The standard deviations are quite small, indicating that it will indeed be feasible to make simple conversions.

TABLE 6-4

Standard deviation of [LR-LR(P.79A)]

(Statistics from 71 telephone set characteristics)

Standard	LR	Band (kHz)	Standard deviation
IEEE	Send Receive Overall	0.3-3.4	0.31 0.1 0.4
OREM-B	Send Receive Overall	0.2-4	0.52 0.38 0.5

6.6 *Conclusions*

A proposed amendment, P.79A, to the loudness rating Recommendation P.79 has been investigated and found adequate for transmission planning purposes. (The algorithm P.79A corresponds in principle to taking a weighted dB average over a log (f)-scale. See § 6.4.4 and [13].

The algorithm P.79A has been shown to give, with sufficient accuracy, numerically equal values with P.79 in typical transmission planning situations. P.79A can be expected, for good reasons, to agree better with subjective values than P.79.

It also seems possible to make simple conversions between P.79A, IEEE and OREM-B loudness ratings.

P.79A has the best additivity accuracy as well as the most simple form.

Administrations are therefore urged to make some comparative studies using this P.79A algorithm in transmission planning.

7 **Information on the Zwicker loudness rating method as used by the French Administration** (Contribution from the French Administration)

7.1 *Introduction*

The method recommended for the evaluation of the quality of a communication in terms of loudness is that of loudness rating (LR). The subjective determination of these equivalents is described in Recommendation P.78, and the objective determination in Recommendation P.79.

However, other parameters (e.g. R25E) have been used, and no universal formula is available to transform these parameters into LR. Therefore it will be useful to have a description of how both values (e.g. R25E and LR) can be objectively measured for a given equipment.

This Section sets out to describe a method for an objective evaluation of loudness losses (R25 equivalents and loudness ratings) used by the French Administration, as stated in the SG XII/3 Report, in Boglarlelle, May 1987 (TD64 revised).

7.2 *Characteristics of the method*

The computation of the loudness of stationary-type signals using the Zwicker algorithm, when applied to the objective measurement of R25 equivalents and LR, gives results which are in good agreement with results obtained using the corresponding subjective evaluation methods [16].

One specific characteristic of this algorithm is that it can be used to compare loudness between systems whose transmitted frequency bandwidth is not within the limits of 300-3400 Hz.

Adaptations of the Zwicker algorithm to monaural and binaural listening modes respectively are used to evaluate different types of terminals: handset telephones [16], operator headsets [17] and loudspeaker telephones [18].

It is essential to use "complex" voices as defined in Recommendation P.51 to measure nonlinear systems. This algorithm has lead to satisfactory results in the evaluation of hand-free sets [19] when using such input voices.

7.3 *Principle of Zwicker's algorithm to calculate loudness*

The method is based on the use of method B of ISO 532 Rec. (Zwicker method).

7.3.1 *Essential phenomena considered in the calculation of loudness*

The algorithm establishes a relationship between the stimulus (physical) and the auditory sensation (psychoacoustical). Loudness is composed of three main phenomena which are as follows: critical bands, masking and equal loudness levels.

7.3.1.1 *Critical bands*

In the human ear, wideband sounds seem to be louder than pure or narrow-band sounds with the same acoustic pressure levels. It is also possible to prove that, around a given frequency and for a fixed level of acoustic pressure, loudness remains constant as long as the sound bandwidth does not exceed a given value called the critical band. If sound bandwidth exceeds this value, a distinct increase in loudness can be noted. In this way, the ear divides the domain of audible frequencies into 24 critical bands, within which any excitation is integrated without weighting. Therefore, loudness measurement is based upon spectral sound analysis.

7.3.1.2 *Masking effect*

The masking effect consists of raising a (reference) sound's threshold level of audibility when transmitting another sound of a (masking) noise with a lower frequency (or higher, but in a weaker proportion). This can be explained via the notion of specific loudness.

7.3.1.2.1 *Specific loudness*

A narrow band of noise or a pure sound, whose spectral energy is concentrated at one specific point of the frequency range, causes a large portion of the basilar membrane (one of the essential auditory organs) to vibrate. This causes not only an excitation of the centre, but also an excitation of the sides which is especially significant in frequency zones above that of the signal frequency. The effect of these side excitations shows itself via the masking effect. The excitations of the centre and the sides contributes to the specific signal loudness.

7.3.1.2.2 *Masked or partially masked loudness*

If a weak sound falls in the highest frequency zone as defined above, it can be masked partially or totally by the side excitations, and therefore does not produce any specific loudness.

7.3.1.3 *Equal loudness levels*

The ear is not equally sensitive to different frequencies; curves of equal loudness levels allow to compare the loudness of sounds produced by sounds of different frequencies once the signal levels have been physically measured. On an "acoustic pressure level vs. frequency" diagram, these curves link the points which correspond to an equal sensation of sound loudness, including those found at the threshold of hearing.

7.3.2 *Global loudness of a complex sound*

Using a spectral analysis in thirds of octaves [1)] the Zwicker method [20] is used to calculate the loudness of stationary signals with the following double correction:

- by transforming the pressure level in each third octave band into a specific loudness,
- by a specific loudness summation, weighted by the masking effect.

7.3.3 *Computer measurement of loudness*

This measurement is possible since the acoustic pressure level for each third octave band is known. The FORTRAN program enabling this loudness to be computed is described in [21].

Note – The Zwicker algorithm includes two variants: one for listening in a free field, the second for listening in a diffuse field. Results referenced in [16], [17] made use of the first variant. Both lead to satisfactory values when applied for the determination of R25 and LR.

7.4 *Application of the Zwicker method for measuring the loudness loss (LL): R25 equivalents (R25E) and loudness ratings (LR)*

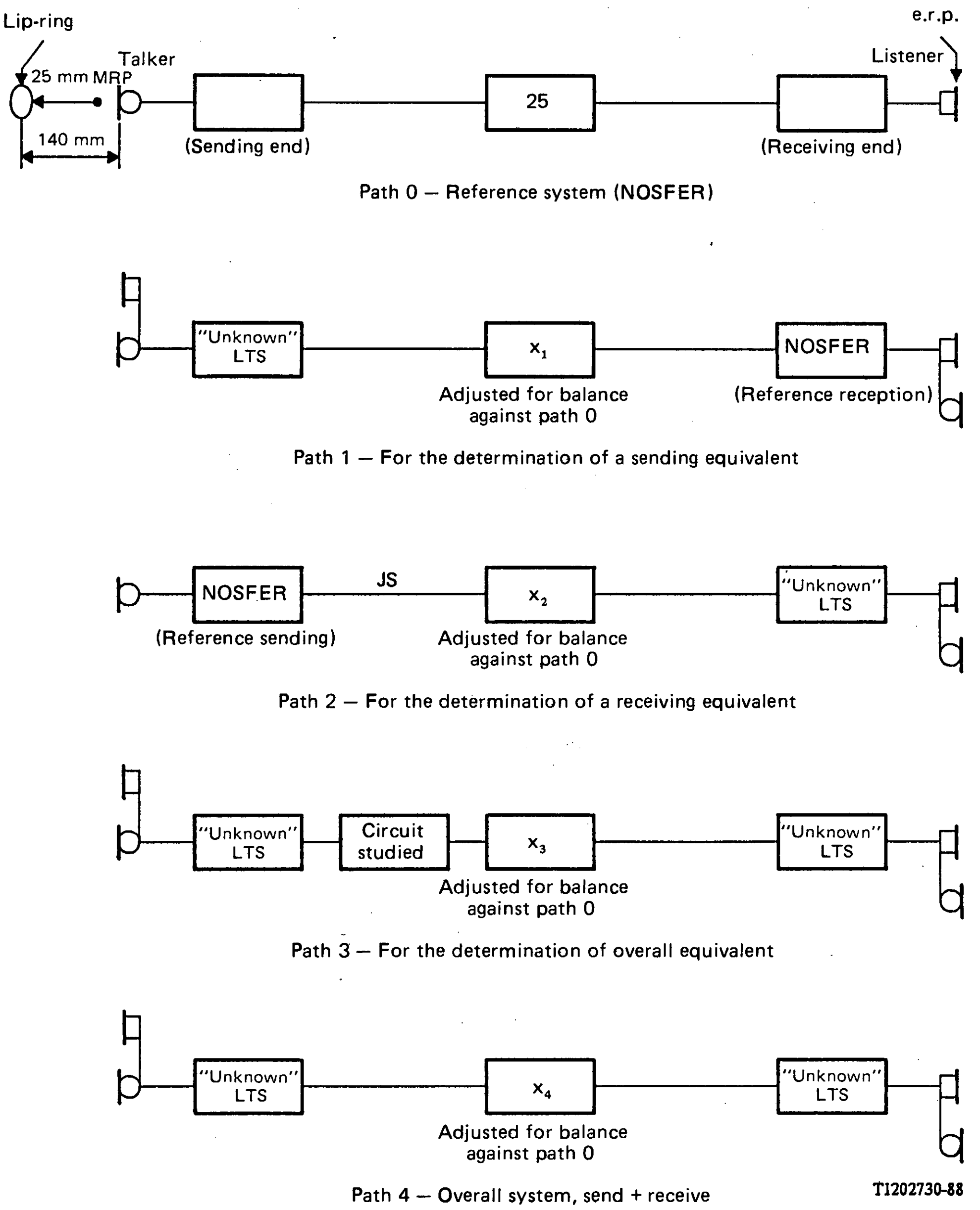
7.4.1 *Fundamental principle of the objective LL measuring method*

The principles on which the following instrumental method are based are similar to the subjective methods described in Figures 7-1 and 7-2 for determining the R25 equivalents and LR. In this method:

a) The speech signals are replaced by an artificial acoustic voice, the spectral characteristics of which are given in Recommendation P.51.
b) The reference parameters and signals are calculated on the basis of the nominal efficiency characteristics, as a function of the frequency of the NOSFER and IRS defined in P.42 *(Red Book)* and P.48. These are:
   - the artificial electric voices resulting from the transmission of acoustic artificial voice via reference transmission systems,
   - the NOSFER reference loudness (RL),
   - the loss x_2 as a result of comparing and equalizing the loudness of NOSFER and IRS paths.
c) The values of "*LE*" relative to acoustic leakage are used as artificial ear/real ear correcting terms.
d) Operator evaluation of sound loudness is replaced by a calculation of the loudness of stationary noise intercepted by a standardized artificial ear, and is performed according to Zwicker's algorithm.

---

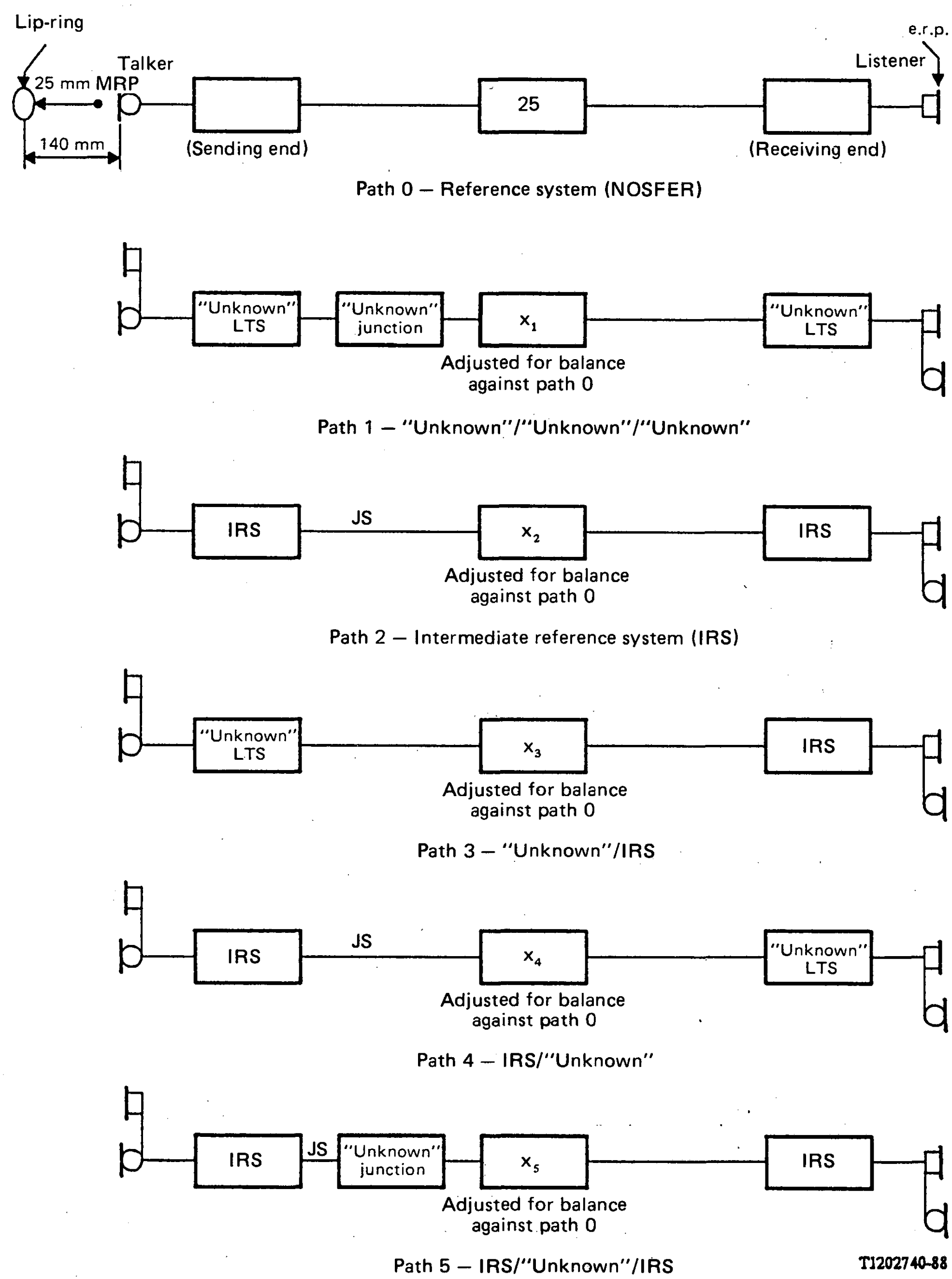
1) Third octave bands are a good approximation of the critical bands, provided that bands lower than 280 Hz are appropriately grouped together.



SR25E $= 25 - x_1$
RR25E $= 25 - x_2$
OSR $= 25 - x_3$
OVE $= 25 - x_4$
LTS Local telephone system

FIGURE 7-1

Arrangement of paths for subjective method of determination of R25 equivalents according to Rec. P.72 (Red Book)



OLR $= x_2 - x_1$
SLR $= x_2 - x_3$
RLR $= x_2 - x_4$
JLR $= x_2 - x_5$

FIGURE 7-2

Arrangement of paths for subjective method of determination of loudness ratings according to Rec. P.78 (Red Book)

7.4.2 *Characteristics of reference signals and parameters*

The reference signals and parameters which objectively characterize speech communication in Figures 7-1 and 7-2 are defined hereafter.

Note – Whether it is a question of human or artificial mouths and ears, the definition of mouth and ear reference points (MRP and ERP) does not change (Annex A of Recommendation P.64) and the standardized speaking positions are identical.

- *Acoustic artificial voice:* defined at the MRP. For spectral characteristics, see Table 2/P.51.
- *Electrical artificial voices:* these voices substitute, at points JS in Figures 7-1 and 7-2, for the human voice/NOSFER or IRS systems. For spectral characteristics see Tables 7-1 and 7-2 (columns N_T and N_S respectively).
- *Reference loudness (RL):* loudness of the acoustic signal at the e.r.p. of "path 0", when the acoustic artificial voice is applied to the MRP.

$$\mathrm{RL} = 21.1 \text{ sones}$$

- x_2: loss calculated according to the flowchart in Figure 7-4 to give equal loudness for paths 0 and 2 (Figure 7-2).

$$x_2 = 21.5 \text{ dB}$$

- *Acoustic leakage "LE":* defined in two cases:
  - i) for the telephone set receiver (Table 4/P.79)
  - ii) for the NOSFER receiver (Table 7-3, LE_R).

TABLE 7-1

Electrical artificial voice of the NOSFER

(output of the NOSFER sending system)

Frequency (Hz)	N_T (dBV)	Frequency (Hz)	N_T (dBV)
100	–29.7	1000	–20.3
125	–24.7	1250	–22.1
160	–21.4	1600	–24.3
200	–19.2	2000	–26.0
250	–18.6	2500	–27.8
315	–18.4	3150	–28.1
400	–18.7	4000	–29.9
500	–18.7	5000	–34.8
630	–18.8	6300	–42.7
800	–19.4	8000	–46.4

TABLE 7-2

Electrical artificial voice of the IRS

(output of the IRS sending system)

Frequency (Hz)	N_S (dBV)	Frequency (Hz)	N_S (dBV)
100	−68.9	1000	−22.6
125	−55.3	1250	−23.3
160	−42.0	1600	−23.6
200	−33.6	2000	−24.8
250	−27.7	2500	−25.1
315	−23.8	3150	−26.8
400	−21.7	4000	−67.0
500	−21.0	5000	−82.8
630	−21.5	6300	−104.5
800	−21.9	8000	−120.5

TABLE 7-3

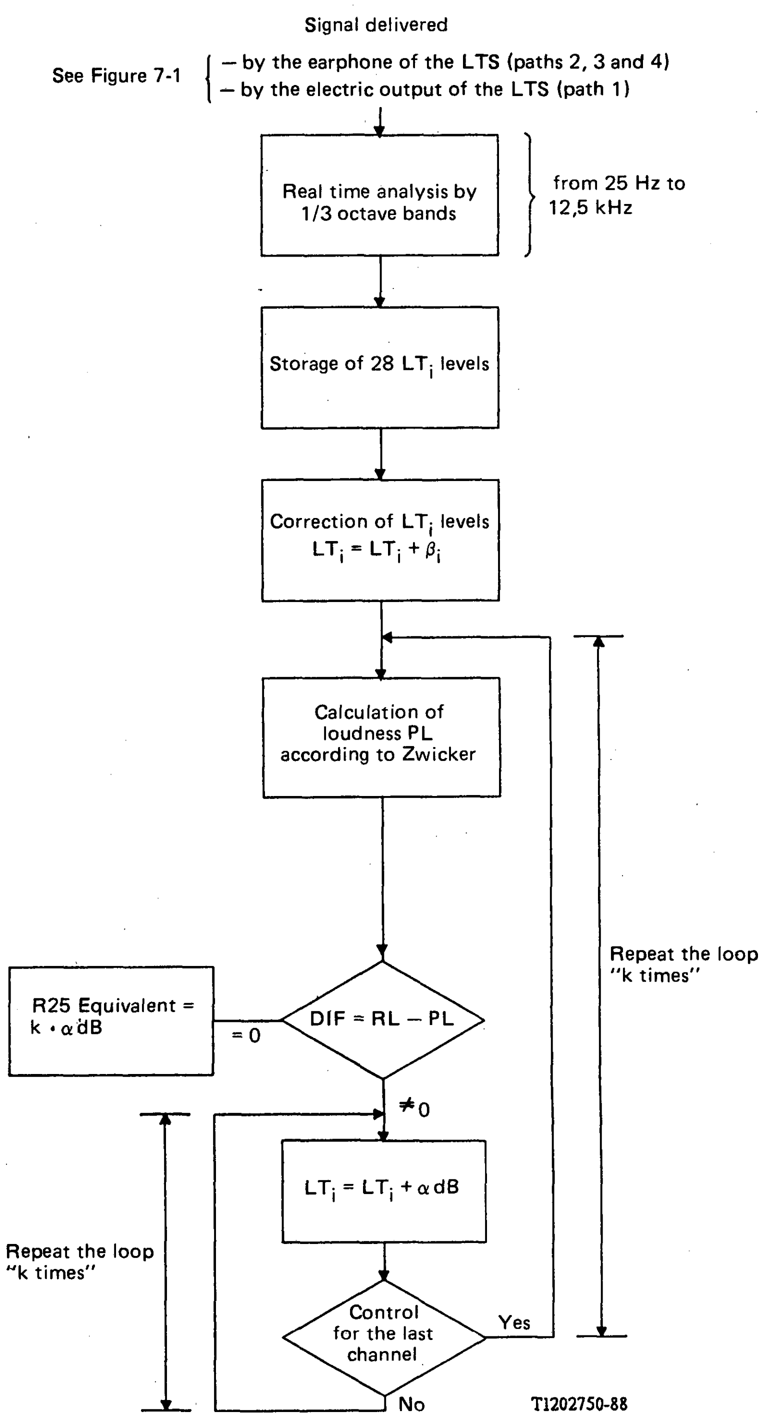
NOSFER receiver acoustical leakage

Frequency (Hz)	LE_R (dB)	Frequency (Hz)	LE_R (dB)
100	0.9	1000	−4.5
125	0.2	1250	−3.9
160	−0.6	1600	−4.6
200	−1.6	2000	−3.3
250	−2.9	2500	−3.2
315	−4.2	3150	−3.3
400	−5.3	4000	−3.7
500	−5.4	5000	−2.9
630	−4.9	6300	−0.8
800	−4.6	8000	−0.8

7.4.3 *Loudness loss computation*

Generally, the measurement consists in comparing and setting the reference loudness (RL) equal to the loudness of the various paths being studied (Figures 7-1 and 7-2). However, considering the structural modification of these paths (see § 7.4.2), the analogy between the subjective and the objective method is only possible if the signals which characterize these paths have been corrected (β_i terms of Figures 7-3 and 7-4: acoustical leakage, nominal sensitivity/frequency of the reference receiving systems) before calculating their loudness and comparing them to the reference.

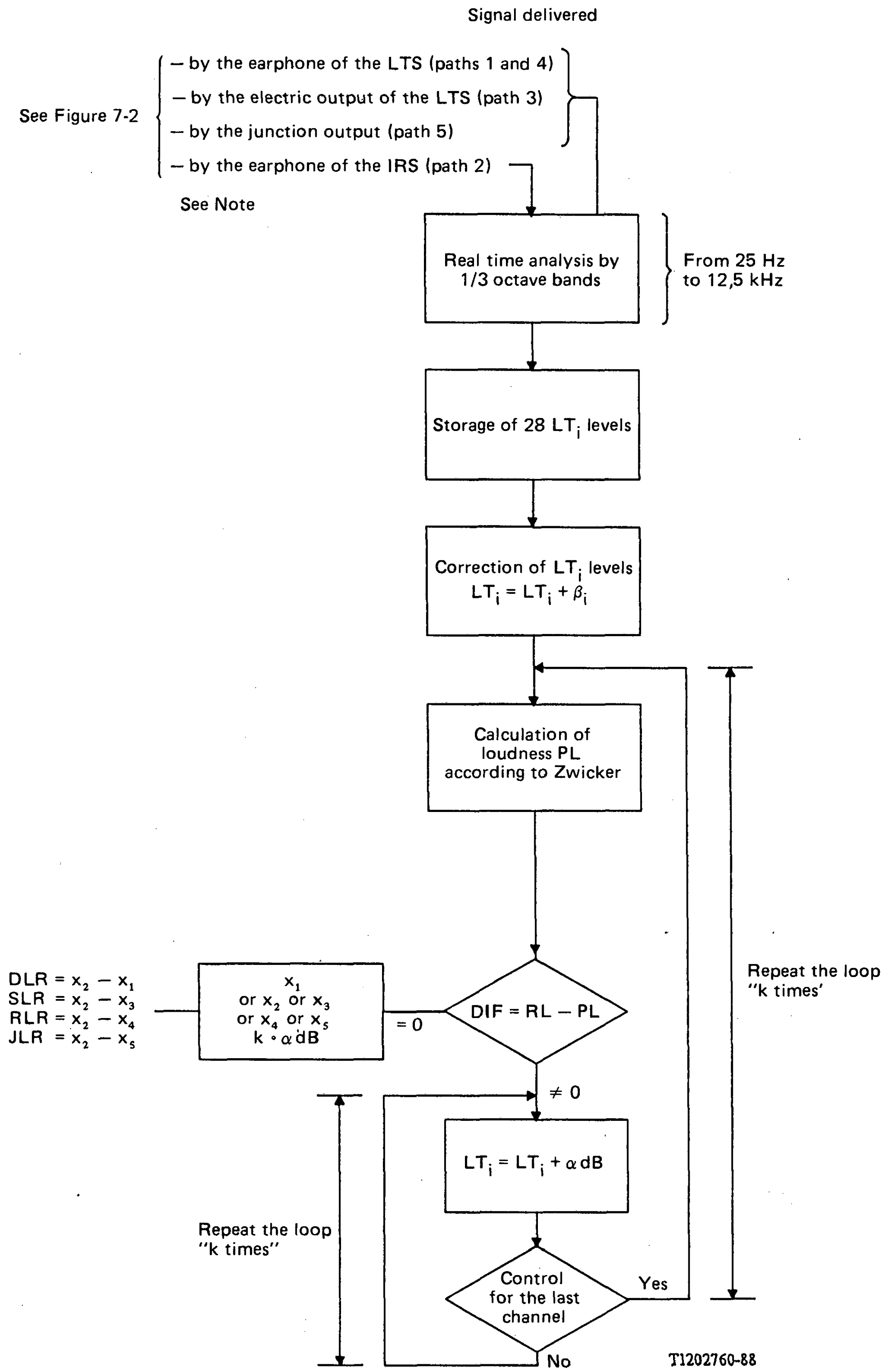
The R25 equivalents and LR are therefore calculated according to the flowcharts in Figures 7-3 and 7-4.



PL: Path loudness under study
RL: Reference loudness
i: Varies from 1 to 28
α dB: Steps of variation in dB of the LT_i levels (positive or negative)
k: number of sequences proportional to the R25 equivalent
β_i: if receiving and overall: leakages LE (Rec. P.79)
if sending: leakages LE_R of NOSFER receiver
sensitivity of the NOSFER system: 25 dB

FIGURE 7-3

Flow chart for the calculation of the R25 equivalents



PL: Path loudness uner study
LR: Reference loudness
i: Varies from 1 to 28
αdB: Steps of variation in dB of LT_i levels (positive or negative)
k: number of sequences proportional to the value of x_i
β_i: if receiving and overall: leakages LE (Rec. P.79)
if sending and junction: leakages LE (Rec. P.79), receiving sensitivities of the IRS

Note – The procedure to calculate x_2 is shown as an example. x_2 = constant = 21.5 dB.

FIGURE 7-4

Flow chart for the calculation of loudness ratings

7.5 *General structure of the CERF apparatus of the French Administration*

The diagram in Figure 7-5 shows how the main elements of the measuring device performing the operation described above are organized. [22] gives a detailed description of the functions and features of the device.

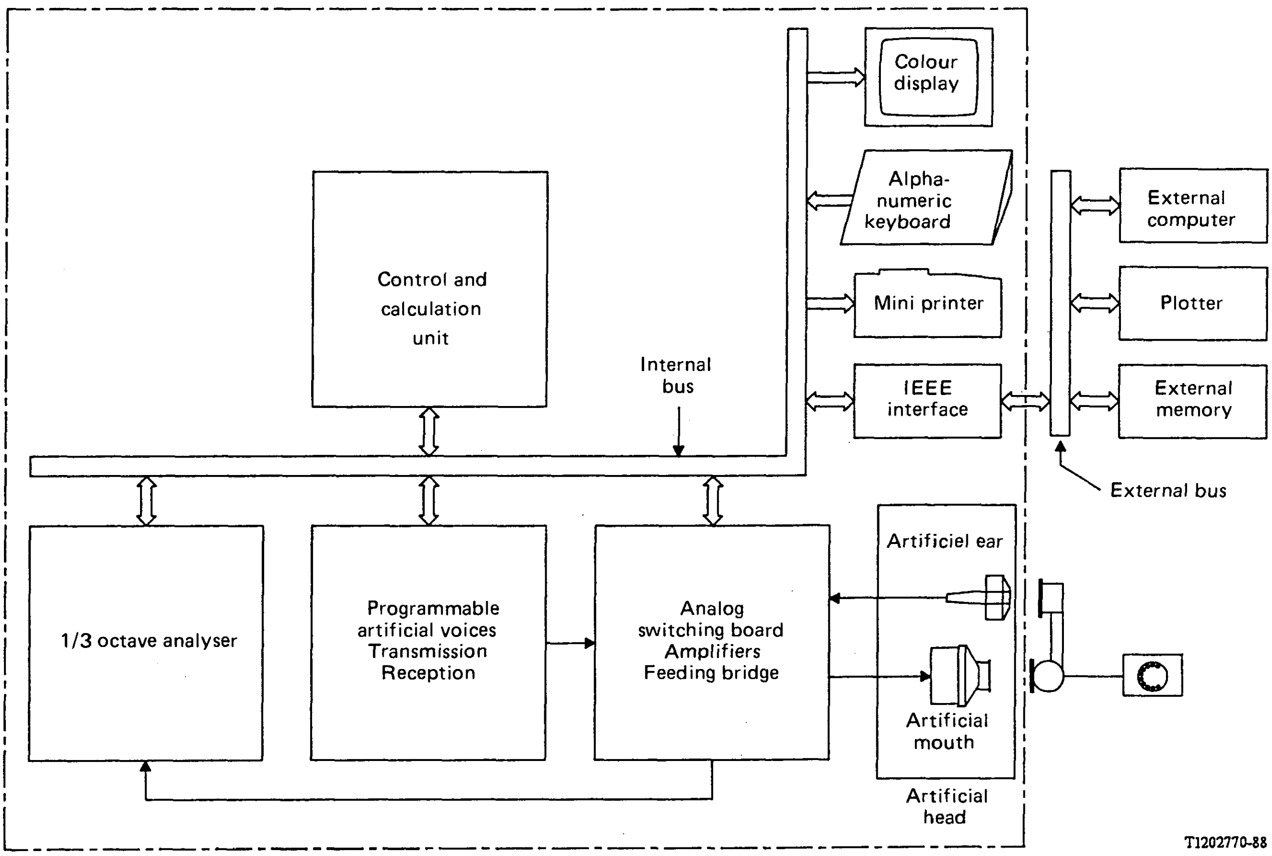


FIGURE 7-5

Block diagram of the CERF system of the French Administration

8 Information on the OREM-B loudness loss method as used by the Administration of the Federal Republic of Germany (Contribution by the Administration of the Federal Republic of Germany)

8.1 *Definition*

Within the area of the Deutsche Bundespost, measurements of loudness ratings are performed according to DIN 44013 "Objective Reference Equivalent Measuring Device OREM-B, Configuration and Application."

8.1.1 *OREM-B loudness related ratings, BD*

By definition, the loudness BD is zero if a sound pressure of 1.6 Pa is reached at the SFERT microphone in the Braun ear at a sound pressure of 1.07 Pa under the measurement conditions specified in DIN 44013.

8.1.2 *OREM-B send loudness, SBD*

The send loudness determined by operating the test item (e.g. a telephone set together with the feeding bridge, possibly also with connected lines and other equipment) as an electric transmitter and by comparing the voltage measured at a 600 ohm terminating impedance with the reference voltage.

By definition, the "SBD" is zero if the output voltage at the SFERT microphone in the presence of a 1.07 Pa sound pressure is 285 mV (see Figure 8-1).

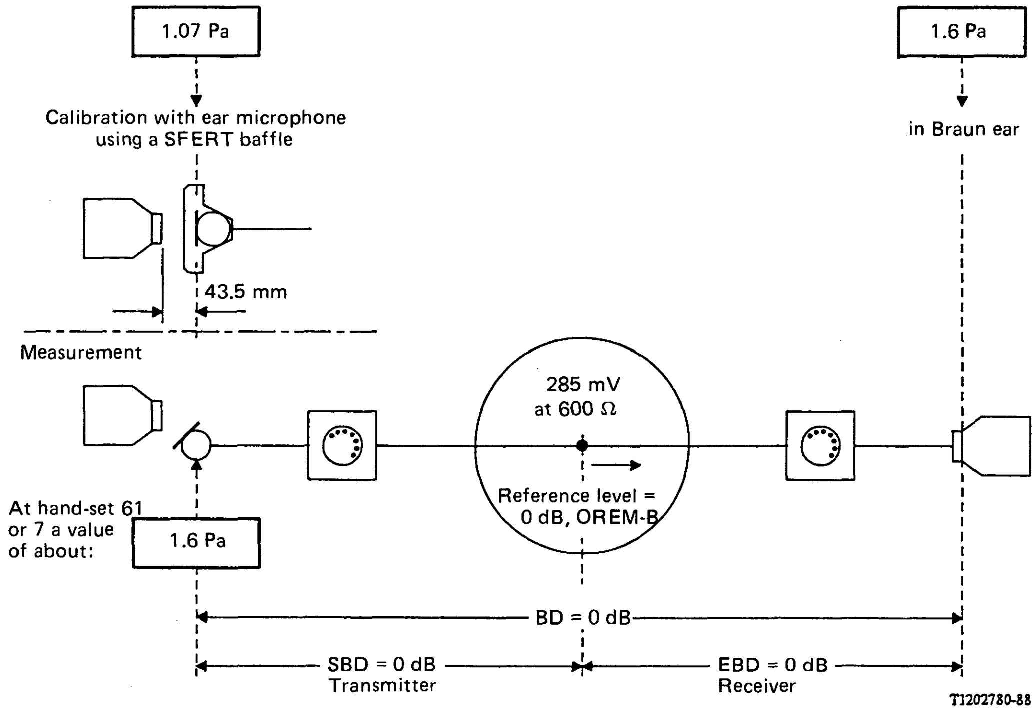


FIGURE 8-1

Block schematic of the OREM-B apparatus

8.1.3 *OREM-B receive loudness, EBD*

The receive loudness determined by operating the test item (e.g. a telephone set together with the feeding bridge, possibly also with connected lines and other equipment) as an electric receiver and by comparing the sound pressure measured in the Braun ear with the reference sound pressure.

By definition, the "EBD" is zero if the sound pressure measured at an open-circuit voltage of the transmitter of 570 mV (internal resistance: 600 ohms) is 1.6 Pa (see Figure 8-1).

8.1.4 *OREM-B overall loudness, OBD*

(Overall loudness (OBD) of a telephone connection): The reference equivalent determined by comparing a complete telephone connection, possibily together with interposed lines and other equipment, with the OREM-B reference transmitter and receiver (see Figure 8-1).

8.1.5 *OREM-B sidetone loudness, RBD*

The sidetone loudness determined by comparing – in transmissions from the microphone to the receiver capsule of the same test item (e.g. a telephone set with a specific terminating impedance) the sound pressure of the receiver with the reference sound pressure.

By definition, it is zero if the sound pressure measured in the Braun ear is 1.6 Pa in the presence of a sound pressure of 1.07 Pa at the SFERT microphone.

8.2 *Measurement conditions deviating from Rec. P.64*

- Instead of the handset position according to Annex A of Rec. P.76 [the loudness rating guard-ring position (LRGP)], a position according to Rec. P.72 [reference equivalent speaking position (RESP) *(Red Book)*] is used.
- Instead of the IEC coupler, a Braun coupler is used.
- Use is made of an artificial mouth according to Rec. P.51.
- Calibration of the artificial mouth is not performed under free-field conditions but with the aid of the SFERT baffle. The sound pressure build-up in the SFERT baffle is compensated by an adequate filter in the generator section. At the diaphragm of the microphone in the SFERT baffle (see Figure 8-2), spaced 43.5 mm apart from the lip plane, the sound pressure is set to 1.07 Pa (sound pressure level: 94.6 dB). Between 200 Hz and 4000 Hz, the sound pressure should be as frequency-independent as possible. In the process, the SFERT filter is activated.
- Via a regulation loop, the sound pressure is kept constant at the calibration value (independently of the test item).

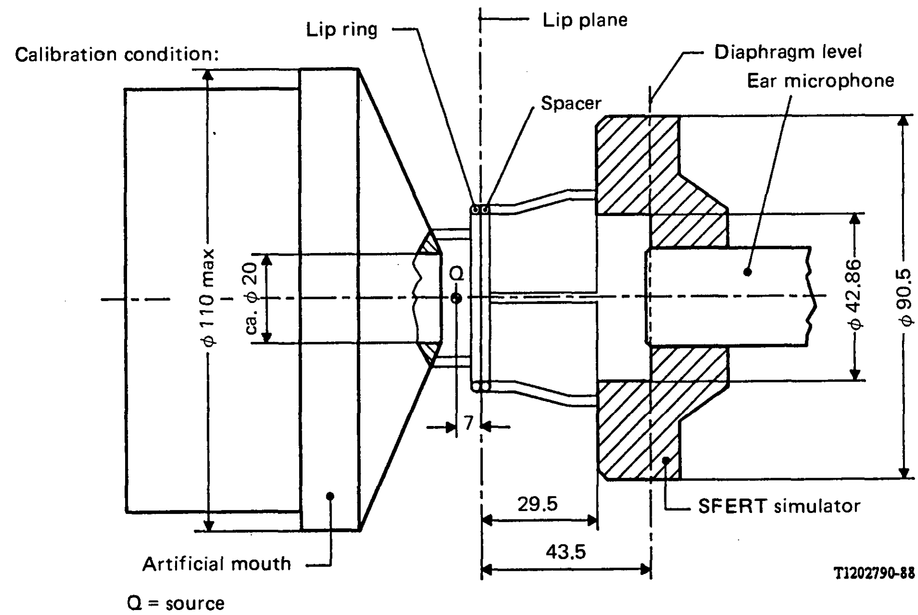


FIGURE 8-2

Artificial mouth with calibration arrangement

8.3 *Algorithm*

The successive voltages U_1, U_2 ... U_n of the swept sinusoidal signal are added according to the following law:

$$U^m = (U_1^m \cdot t_1 + U_2^m \cdot t_2 + U_3^m \cdot t_3 + \dots U_n^m \cdot t_n) \cdot \frac{1}{\sum t_i}$$

for t_i approching 0:

$$U^m = (1/T) \int_0^T [U(t)]^m \, dt$$

The exponent m is 0.6.

The static transient time of the indicator is 3.5 s.

The frequency sweep 200 ... 4000 ... 200 Hz is logarithmic with time, with a complete sweep cycle per second.

ANNEX A

(to Supplement 19 – ref. to § 5.3)

E-factor coefficients

The attenuation values L_i are given for the "wideband" $f_1 \ldots f_N$. The "wideband" $LR(W)$ is computed using the algorithm:

$$\left.\begin{aligned} LR(W) &= L_0(W) + L_m(W) \\ L_m(W) &= -(1/b)\log_{10}\sum_{i=1}^{N} K_{wi}\,10^{-bL_i} \end{aligned}\right\} \tag{A-1}$$

For shortness we use the notation:

$$b = m/10 \tag{A-2}$$

The "common" band LR is computed in the narrower range:

$$f_{N1} \ldots f_{N2} \qquad f_{N1} > f_1;\, f_{N2} < f_N$$

$$\left.\begin{aligned} LR &= L_0 + L_m \\ L_m &= -(1/b)\log_{10}\sum_{i=N1}^{N2} K_i\,10^{-bL_i} \end{aligned}\right\} \tag{A-3}$$

According to the definition of the K_i-coefficients we have:

$$\left.\begin{aligned} \sum_{i=1}^{N} K_{wi} &= 1 \\ \sum_{i=N1}^{N2} K_i &= 1 \end{aligned}\right\} \tag{A-4}$$

The relationship between the two algorithms is defined to be as follows:

For a strictly band-limited system, i.e. one for which $L_1 \ldots L_{N1} = \infty$, we let:

$$LR(W) = LR \tag{A-5}$$

and further we set:

$$K_{wi}\,/\,K_i = D;\ \text{a constant.} \tag{A-6}$$

Thus, we get:

$$D = \sum_{i=N1}^{N2} K_{wi} \tag{A-7}$$

and

$$L_0 = L_0(W) - (1/b)\log_{10} D \tag{A-8}$$

We will use the following notation:

$$\Sigma' = \sum_{i=1}^{N1-1} + \sum_{i=N2+1}^{N} \tag{A-9}$$

In the general case we have:

$$LR(W) = L_0(W) - (1/b)\log_{10}\left\{\sum_{i=N1}^{N2} K_{wi}\,10^{-bL_i} + \Sigma' K_{wi}\,10^{-bL_i}\right\} \tag{A-10}$$

which results in

$$LR(W) = LR - (1/b)\log_{10}\left\{1 + \Sigma'(K_{wi}/D)\,10^{-bL_i}\right\} \tag{A-11}$$

As the terms in the sum Σ' in Equation (A-11) are small, one can make a series expansion. Thus:

$$LR(W) = LR - \Sigma' C_i\,10^{-bL_i} \tag{A-12}$$

where

$$C_i = K_{wi}/(bD \ln 10) \tag{A-13}$$

The last term in Equation (A-12) is designated as the loudness improvement, the *E*-factor.

$$E = \Sigma' C_i\,10^{-bL_i} \tag{A-14}$$

In the actual case the "wideband" can be taken to encompass the range $f_1 = 200$ Hz to $f_{14} = 4000$ Hz and the "common" band $f_3 = 315$ Hz to $f_{13} = 3150$ Hz. The "band edge" frequencies are 200, 250 and 4000 Hz.

The coefficients C_i have been computed for some LR algorithms under discussion using the K_i-values for OLR as given in Table A-1. (Details of the algorithms are given in [11] and [7] as well as the method used for converting W_i-weights to K_i-weights.) The C_i-values are presented in Table A-2.

The P.79 algorithm or its smoothed version P.79/S are not suitable for band edge performance, analysis as their frequency weighting has been shown to be less correct (too much emphasis on the lower frequencies [7].)

It is apparent from Table A-2 that the C_i-coefficients from the three algorithms do not differ very much. As the human speech and hearing characteristics at the band edges can be expected to vary rather much, the actual values of the C_i's cannot be critical. Therefore, it is reasonable to use the "rounded-off" values:

$$C(0.2) = 0.5; \qquad C(0.25) = 1; \qquad C(4) = 1 \text{ dB} \tag{A-15}$$

and to set $m = 0.2$.

Which algorithm should be used when computing the "common"-band LR? As shown in [11] and [7], there are several algorithms which correlate about equally well with subjective measurements. The simplest one is the algorithm "C" which was therefore chosen.

TABLE A-1

Algorithm	m	K_{wi}		
		0.2 kHz	0.25 kHz	4 kHz
P.XXE	0.225	0.0227	0.0389	0.0292
CH	0.2	0.0306	0.0439	0.0324
B	0.2	0.031	0.042	0.042
P.79/S	0.175	0.0536	0.0765	0.0243

TABLE A-2

Algorithm	C_i			$\sum C_i$
	0.2 kHz	0.25 kHz	4 kHz	
P.XXE	0.48	0.83	0.62	1.93
CH	0.74	1.06	0.79	2.59
B	0.76	1.03	1.03	2.82
Mean	0.66	0.97	0.81	2.45

ANNEX B

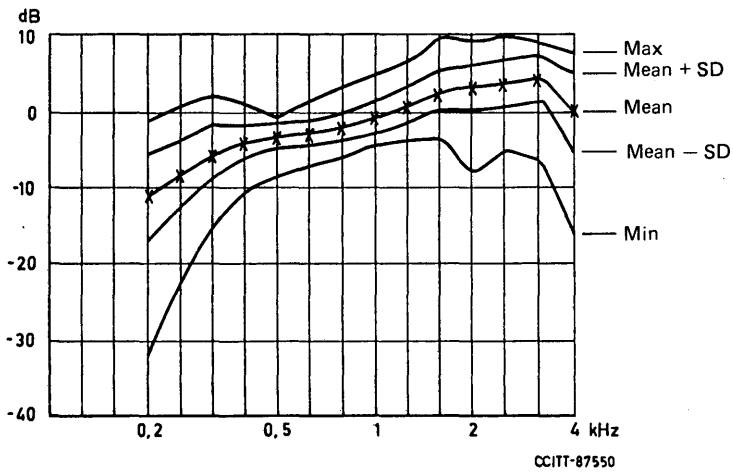
(to Supplement 19 – ref. to § 6.3)

Seventy-one telephone set characteristics were obtained of which:

a) thirteen from CCITT COM XII-No. 164 (1977-1980), and

b) fifty-eight from Barnes in a private communication.

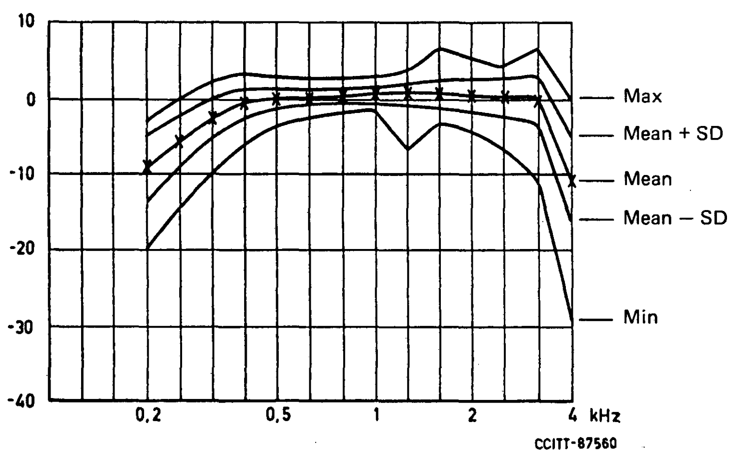
The statistics of the send and receive sensitivity curves are shown in Figures B-1 and B-2. The curves were normalized by subtracting the "average" sensitivity computed by Equation (6-1b).



SD standard deviation

FIGURE B-1

Statistics for the normalized send curves



SD standard deviation

FIGURE B-2

Statistics for the normalized receive curves

References

[1] IEEE Standard 661-1979, *Standard method for determining objective loudness ratings of telephone connections, IEEE*, New York.

[2] CCITT Draft Recommendation P.XXE, Question 15/XII, Annex 2 (II) Contribution COM XII-No. 1, Study Period 1977-1980, Geneva, 1976

[3] CCITT Contribution COM XII-No. 160, *Use of Zwicker's method for computing loudness ratings* (Australia), Study Period 1981-1984, Geneva, 1983.

[4] BARNES (J.): Private Communication, STL.

[5] CCITT Contribution COM XII-No. 10 (People's Republic of China), Study Period 1981-1984.

[6] CCITT Contribution COM XII-No. 55 (ELLEMTEL), Study Period 1981-1984.

[7] CCITT Contribution COM XII-No. 194 (ELLEMTEL), Study Period 1981-1984.

[8] CCITT Contribution COM XII-No. 98, *Calculation of loudness ratings for different bandwidths* (British Telecom), Study Period 1981-1984, Geneva, 1982.

[9] CCITT Contribution COM XII-No. 78, *Additivity of loudness ratings* (ELLEMTEL), Study Period 1981-1984, Geneva, 1982.

[10] CCITT Contribution COM XII-No. 111, *Choice of representative loss/frequency characteristics for electrical elements* (ITT/STL), Study Period 1981-1984, Geneva, 1983.

[11] CCITT Contribution COM XII-No. 176, *Loudness rating algorithms applied to some published data* (ELLEMTEL), Study Period 1981-1984, Geneva, 1984.

[12] CCITT Contribution COM XII-No. 4, *The stability of the CCITT test team.* Some comments. (ELLEMTEL), Study Period 1985-1988, Geneva, 1985.

[13] CCITT Contribution COM XII-No. 1, Question 15/XII (Annex 2), *Adapting loudness ratings for transmission planning*, Study Period 1985-1988.

[14] Private Communication from FTZ (Federal Republic of Germany).

[15] CCITT Contribution COM XII-No. 15, *Objective measurement and calculation of the reference attenuation of local cables* (Sweden), Study Period 1964-1968.

[16] CCITT Contribution COM XII-No. 111, *Subjective and objective comparison of results concerning loudness ratings (LR) and R25 equivalents (R25E)*, Study Period 1985-1988.

[17] CCITT Contribution COM XII-No. 58, *Determination of loudness ratings and R25 equivalents for operator headsets, taking acoustical leakage into account*, Study Period 1985-1988.

[18] CCITT Contribution COM XII-D.44 – WPXII/2 – Shangaï, 12-16 October 1987 Status Report of Question 17: "Loudspeaking telephones" (see COM XII-R 20, Study Period 1985-1988).

[19] CCITT Contribution COM XII-235, *Sending objective measurements of hands-free sets. Influence of artificial voices characteristics*, Study Period 1985-1988.

[20] ZWICKER, FELDTKELLER: Das Ohr als Nachrichtenempfänger, *S. Hirzel Verlag*, Stuttgart.

[21] PAULUS (E.) and ZWICKER (E.): Computer programmes for calculating loudness from third-octave band levels or from critical band levels, *Acustica*, 1972, Vol. 27, pp. 253-266.

[22] CCITT Handbook on telephonometry (CERF: an equipment for the objective measurement of various sorts of loudness used by the French Postal and Telecommunication Administration), ITU, Geneva, 1987.

ISBN 92-61-03441-1