INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

G.722.2

Annex B

(01/2002)

SERIES G: TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

Digital terminal equipments – Coding of analogue signals by methods other than PCM

Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)

Annex B: Source Controlled Rate operation

ITU-T Recommendation G.722.2 – Annex B

ITU-T G-SERIES RECOMMENDATIONS

TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

INTERNATIONAL TELEPHONE CONNECTIONS AND CIRCUITS	G.100–G.199
GENERAL CHARACTERISTICS COMMON TO ALL ANALOGUE CARRIER-TRANSMISSION SYSTEMS	G.200–G.299
INDIVIDUAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON METALLIC LINES	G.300–G.399
GENERAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON RADIO-RELAY OR SATELLITE LINKS AND INTERCONNECTION WITH METALLIC LINES	G.400–G.449
COORDINATION OF RADIOTELEPHONY AND LINE TELEPHONY	G.450–G.499
TESTING EQUIPMENTS	G.500–G.599
TRANSMISSION MEDIA CHARACTERISTICS	G.600–G.699
DIGITAL TERMINAL EQUIPMENTS	G.700–G.799
General	G.700–G.709
Coding of analogue signals by pulse code modulation	G.710–G.719
Coding of analogue signals by methods other than PCM	**G.720–G.729**
Principal characteristics of primary multiplex equipment	G.730–G.739
Principal characteristics of second order multiplex equipment	G.740–G.749
Principal characteristics of higher order multiplex equipment	G.750–G.759
Principal characteristics of transcoder and digital multiplication equipment	G.760–G.769
Operations, administration and maintenance features of transmission equipment	G.770–G.779
Principal characteristics of multiplexing equipment for the synchronous digital hierarchy	G.780–G.789
Other terminal equipment	G.790–G.799
DIGITAL NETWORKS	G.800–G.899
DIGITAL SECTIONS AND DIGITAL LINE SYSTEM	G.900–G.999
QUALITY OF SERVICE AND PERFORMANCE	G.1000–G.1999
TRANSMISSION MEDIA CHARACTERISTICS	G.6000–G.6999
General	G.6000–G.6099
Symmetric cable pairs	G.6100–G.6199
Land coaxial cable pairs	G.6200–G.6299
Submarine cables	G.6300–G.6499
Optical fibre cables	G.6500–G.6599
Characteristics of optical components and subsystems	G.6600–G.6999
DIGITAL TERMINAL EQUIPMENTS	G.7000–G.7999
General	G.7000–G.7099
Coding of analogue signals by pulse code modulation	G.7100–G.7199
Coding of analogue signals by methods other than PCM	G.7200–G.7299
Principal characteristics of primary multiplex equipment	G.7300–G.7399
Principal characteristics of second order multiplex equipment	G.7400–G.7499
Principal characteristics of higher order multiplex equipment	G.7500–G.7599
Principal characteristics of transcoder and digital multiplication equipment	G.7600–G.7699
Operations, administration and maintenance features of transmission equipment	G.7700–G.7799
Principal characteristics of multiplexing equipment for the synchronous digital hierarchy	G.7800–G.7899
Other terminal equipment	G.7900–G.7999
DIGITAL NETWORKS	G.8000–G.8999
General aspects	G.8000–G.8099
Design objectives for digital networks	G.8100–G.8199
Quality and availability targets	G.8200–G.8299
Network capabilities and functions	G.8300–G.8399
SDH network characteristics	G.8400–G.8499
Management of transport network	G.8500–G.8599
SDH radio and satellite systems integration	G.8600–G.8699
Optical transport networks	G.8700–G.8799

For further details, please refer to the list of ITU-T Recommendations.

ITU-T Recommendation G.722.2

Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)

Annex B

Source Controlled Rate operation

Summary

This annex describes the operation of the adaptive multi-rate wideband speech codec during Source Controlled Rate (SCR) operation.

The SCR operation described here was also adopted by 3GPP in 3GPP specification TS 26.193.

Source

Annex B to ITU-T Recommendation G.722.2 was prepared by ITU-T Study Group 16 (2001-2004) and approved under the WTSA Resolution 1 procedure on 13 January 2002.

FOREWORD

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

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

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

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementors are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database.

CONTENTS

ITU-T Recommendation G.722.2

Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)

Annex B

Source Controlled Rate operation

B.1 Scope

This annex describes the Source Controlled Rate (SCR) operation (also called Discontinuous Transmission, DTX) of the Adaptive Multi-Rate Wideband speech codec. Implementation of this annex is necessary for interoperability with 3GPP systems, but its use is not limited to mobile applications.

The description is structured according to the block diagram in Figure B.1. This structure of distributing the various functions between system entities is not mandatory for implementation, as long as the operation on the speech decoder output remains the same.

B.2 Definitions, symbols and abbreviations

B.2.1 Definitions

This annex, defines the following terms.

B.2.1.1 frame: Time interval of 20 ms, corresponding to the time segmentation of the Adaptive Multi-Rate Wideband speech Codec, also used as a short term for a traffic frame.

B.2.1.2 traffic frame: Block of 132..477 information bits transmitted on the speech traffic channels.

B.2.1.3 SID frame: Frame that conveys information about the acoustic background noise.

B.2.1.4 speech frame: Traffic frame that has been classified as SPEECH_GOOD, SPEECH_BAD or SPEECH_LOST frame.

B.2.1.5 VAD flag: Boolean flag, generated by the VAD algorithm indicating the presence ("1") or the absence ("0") of a speech frame.

B.2.1.6 RX_TYPE: Classifies the received frame.

B.2.1.7 TX_TYPE: Classifies the frame to be transmitted.

B.2.1.8 hangover period: A period of frames added at the end of a speech burst in which VAD flag = "0" and TX_TYPE is = "SPEECH_GOOD", this period provides the encoder with an extra window to derive the Comfort Noise parameters.

B.2.2 Symbols

This annex uses the following symbol:

$N_{elapsed}$ Number of elapsed frames since the last updated SID frame.

B.2.3 Abbreviations

This annex uses the following abbreviations:

AN Access Network

RAN Radio Access Network

RX Receive

SCR Source Controlled Rate (operation)

SID Silence Insertion Descriptor

TS Telecommunication Standard, Technical Specification

TX Transmit

UE User Equipment

VAD Voice Activity Detector

B.3 General

Source Controlled Rate operation (SCR) is a mechanism for the AMR Wideband Speech codec, which allows to encode the input signal at a lower average rate by taking speech inactivity into account. The SCR scheme may be used for the following purposes:

– to save power in the User Equipment;

– to reduce the overall interference and load in the networks.

SCR in the transmitting path (uplink) shall be in operation in UEs, if commanded so by the network. The UE shall handle SCR in the receiving path (downlink) at any time, regardless whether SCR in the transmitting path is commanded or not.

B.3.1 General organisation

The default SCR mechanism described in this annex requires the following functions:

– a Voice Activity Detector (VAD) on the transmit (TX) side;

– evaluation of the background acoustic noise on the transmit (TX) side, in order to transmit characteristic parameters to the receive (RX) side;

– generation on the receive (RX) side of a similar noise, called comfort noise, during periods where the transmission is switched off.

The Voice Activity Detector (VAD) is defined in the main body of ITU-T Rec. G.722.2 and the comfort noise functions in Annex A/G.722.2. Both are based partly on the speech Codec and its internal variables, defined in the main body of ITU-T Rec. G.722.2.

In addition to these functions, if the parameters arriving at the RX side are detected to be seriously corrupted by errors, the speech or comfort noise must be generated from substituted data in order to avoid seriously annoying effects for the listener. These functions are defined in Appendix I/G.722.2.

An overview of one link SCR operation is shown in Figure B.1.

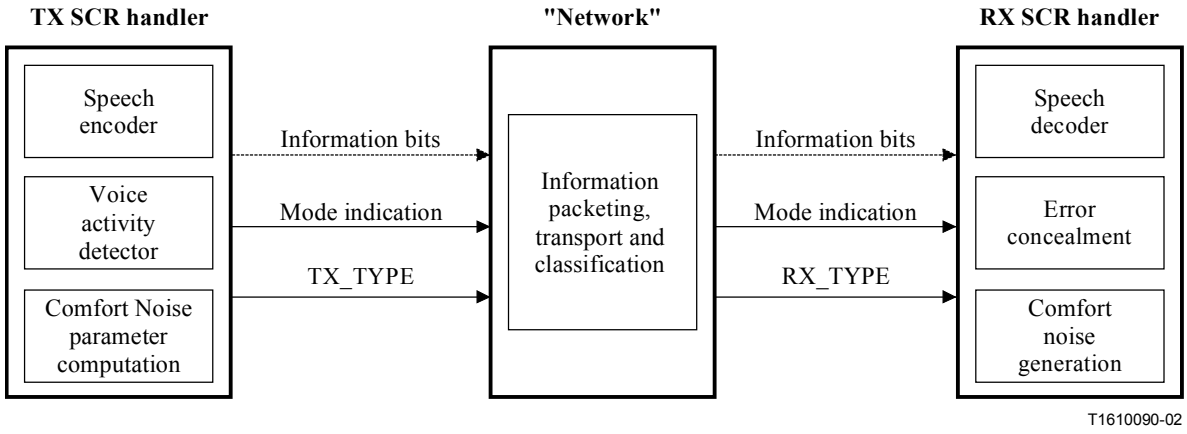


Figure B.1/G.722.2 – Block diagram of one link SCR operation

B.4 AMR-WB SCR operation

B.4.1 Transmit (TX) side

A block diagram of the transmit side SCR functions is shown in Figure B.2.

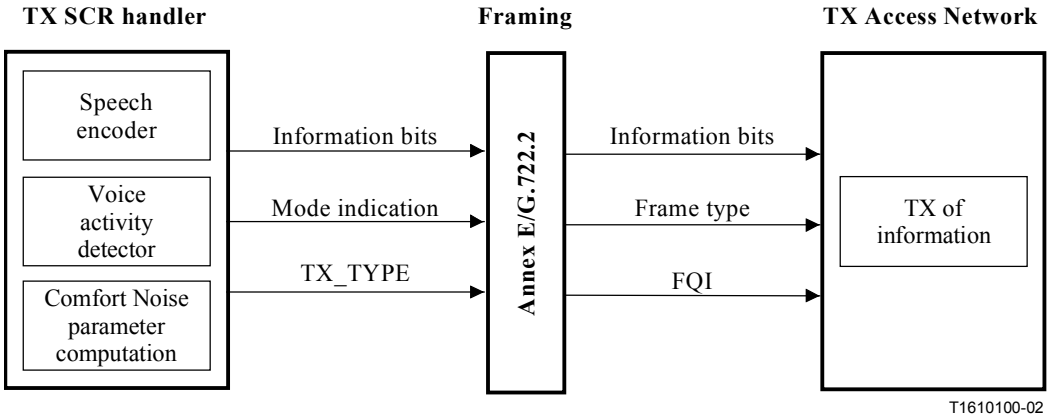


Figure B.2/G.722.2 – Block diagram of SCR functions at the TX side

B.4.1.1 General operation

The TX SCR handler passes traffic frames, individually marked by TX_TYPE, to the Framing unit. Each frame consists of bit fields containing the information bits, the codec mode indication, and the TX_TYPE. TX_TYPE shall be used to specify the contents of the frame. Table B.1 provides an overview of the different TX_TYPEs used and explains the required contents in the information bit and the mode indication bit fields.

Table B.1/G.722.2 – SCR TX_TYPE identifiers

TX_TYPE	Information bits	Mode indication
SPEECH_GOOD	Speech frame, size 132..477 bits, depending on codec mode	Current codec mode
SPEECH_BAD	Corrupt speech frame (bad CRC), size 132..477 bits, depending on codec mode	Current codec mode
SPEECH_LOST	No useful information (NOTE – If implementation does not support the SPEECH_LOST, SPEECH_BAD shall be used instead.)	No useful information
SID_FIRST	Marker for the end of talkspurt, no further information, all 35 comfort noise bits set to "0"	The codec mode that would have been used if TX_TYPE had been "SPEECH_GOOD"
SID_UPDATE	35 comfort noise bits	The codec mode that would have been used if TX_TYPE had been "SPEECH_GOOD"
SID_BAD	Corrupt SID update frame (bad CRC)	The codec mode that would have been used if TX_TYPE had been "SPEECH_GOOD"
NO_DATA	No useful information, nothing to be transmitted	No useful information

TX_TYPE = "SPEECH_LOST" indicates that the Information Bit and Codec Mode fields do not contain any useful data (but still should be transmitted over AN). The purpose of this TX_TYPE is to indicate that the frame was transmitted but lost on some previous phase. This TX_TYPE may occur only in Tandem Free Operation and Transcoder Free Operation situations. Note that it is possible to replace SPEECH_LOST with SPEECH_BAD but this may degrade the quality of the error concealment in the receiving end because concealment may try to use part of the received parameters from the frame which do not contain any useful information.

TX_TYPE = "NO_DATA" indicates that the Information Bit and Codec Mode fields do not contain any useful data (and should not be transmitted over AN). The purpose of this TX_TYPE is to provide the option to save network transmission between the transcoder and AN.

Note that the TX_TYPEs "SPEECH_BAD", "SPEECH_LOST" and "SID_BAD" may occur in Tandem Free Operation and Transcoder Free Operation situations.

The scheduling of the frames for transmission on the Access Network is controlled by the TX SCR handler by the use of the TX_TYPE field.

B.4.1.2 Functions of the TX SCR handler

If TX SCR operation is disabled, the TX SCR handler continuously generates speech frames, i.e. frames marked with TX_TYPE = "SPEECH_GOOD".

If the TX SCR operation is enabled, the VAD flag controls the TX SCR handler operation as described in the following paragraphs.

B.4.1.2.1 AMR-WB SCR timing procedures

To allow an exact verification of the TX SCR handler functions, all frames before the reset of the system are treated as if there were speech frames of an infinitely long time. Therefore, and in order to ensure the correct estimation of comfort noise parameters at RX SCR side, the first 7 frames after the reset or after enabling the SCR operation shall always be marked with TX_TYPE = "SPEECH_GOOD", even if VAD flag = "0" (hangover period: see Figure B.3).

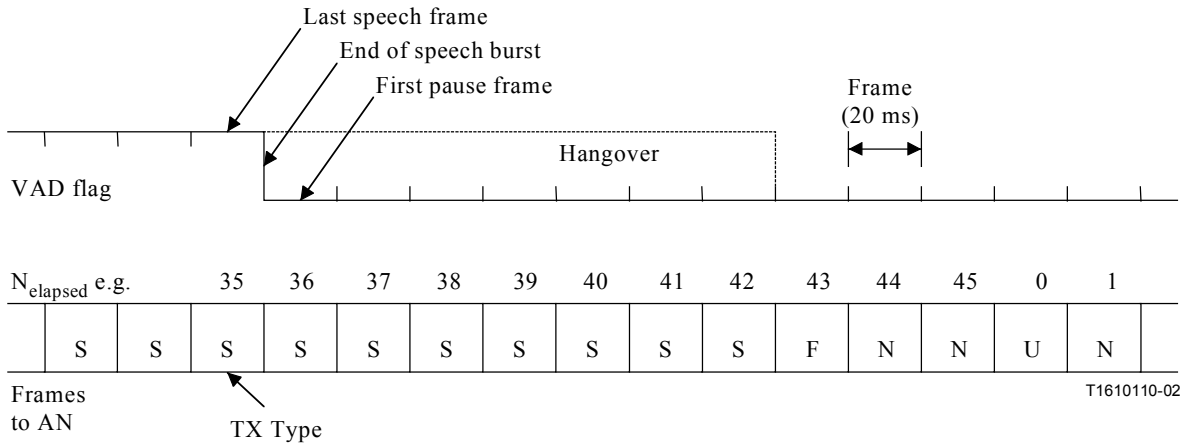
The Voice Activity Detector (VAD) shall operate all the time in order to assess whether the input signal contains speech or not. The output is a binary flag (VAD flag = "1" or VAD flag = "0", respectively) on a frame by frame basis (see the main body of ITU-T Rec. G.722.2).

The VAD flag controls indirectly, via the TX SCR handler operations described below, the overall SCR operation on the transmit side.

Whenever VAD flag = "1", the speech encoder output frame along with mode information shall be passed directly to the AN, marked with TX_TYPE = " SPEECH_GOOD".

At the end of a speech burst (transition VAD flag = "1" to VAD flag = "0"), it takes eight consecutive frames to make a new updated SID analysis available (see Annex A/G.722.2). Normally, the first seven speech encoder output frames after the end of the speech burst shall therefore be passed directly to the AN, marked with TX_TYPE = "SPEECH_GOOD" ("hangover period").

The end of the speech is then indicated by passing frame eight after the end of the speech burst to the AN, marked with TX_TYPE = "SID_FIRST" (see Figure B.3). SID_FIRST frames do not contain data.



TX Types S = SPEECH; F = SID_FIRST; U = SID_UPDATE; N = NO DATA

$N_{elapsed}$ Number of elapsed frames since last SID_UPDATE

Figure B.3/G.722.2 – Normal hangover procedure for AMR-WB ($N_{elapsed}$ > 23)

If, however, at the end of the speech burst, less than 24 frames have elapsed since the last SID_UPDATE frame was computed, then this last analysed SID_UPDATE frame should be passed to the AN whenever a SID_UPDATE frame is to be produced, until a new updated SID analysis is available (8 consecutive frames marked with VAD flag = "0").This reduces the load on the network in cases where short background noise spikes are taken for speech, by avoiding the "hangover" waiting for the SID frame computation.

Once the SID_FIRST frame has been passed to the AN, the TX SCR handler shall at regular intervals compute and pass updated SID_UPDATE (Comfort Noise) frames to the AN as long as VAD flag = "0". SID_UPDATE frames shall be generated every 8th frame. The first SID_UPDATE shall be sent as the third frame after the SID_FIRST frame.

The speech encoder is operated in full speech modality if TX_TYPE = "SPEECH_GOOD" and otherwise in a simplified mode, because not all encoder functions are required for the evaluation of comfort noise parameters and because comfort noise parameters are only to be generated at certain times.

B.4.1.3 The TX part of the AN

The TX part of the AN has the following overall functionality. The transmission is cut after the transmission of a SID_FIRST frame when the speaker stops talking. During speech pauses the transmission is resumed at regular intervals for transmission of one SID_UPDATE frame, in order to update the generated comfort noise on the RX side. The operation of the TX part of the AN is controlled by the TX SCR handler via the TX_TYPE.

All frames, marked with SPEECH_GOOD, SID_FIRST or SID_UPDATE shall be transmitted by the TX part of the AN.

B.4.2 Receive (RX) side

A block diagram of the receive side SCR functions is shown in Figure B.4.

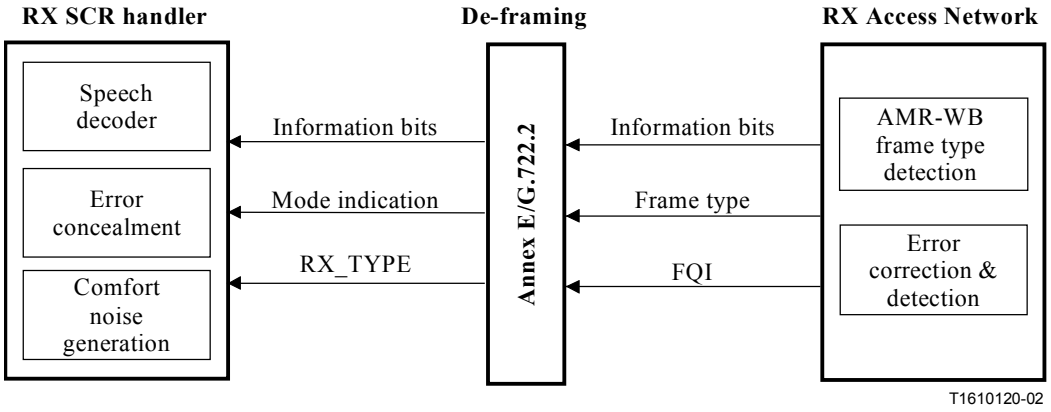


Figure B.4/G.722.2 – Block diagram of the receive side SCR functions

B.4.2.1 General operation

The AN passes all the received traffic frames to the RX SCR handler, classified with RX_TYPE, as described in Table B.2 (see Annex E/G.722.2). The RX SCR handles the frame accordingly.

Table B.2/G.722.2 – RX_TYPE identifiers for AMR-WB

RX_TYPE	Information bits
SPEECH_GOOD	Speech frame without detected errors
SPEECH_BAD	(Likely) speech frame with bad CRC (or estimated to be very bad by the RX part of the AN)
SPEECH_LOST	No frame received. Indicates that this frame was transmitted, but never received.
SID_FIRST	This SID-frame marks the beginning of a comfort noise period.
SID_UPDATE	Correct SID update frame
SID_BAD	Corrupt SID update frame (bad CRC; applicable only for SID_UPDATE frames)
NO_DATA	Nothing useable was received. The synthesis mode of the previous frame type is used.

B.4.2.3 Demands on the RX SCR handler

The RX SCR handler is responsible for the overall SCR operation on the RX side. It consists of two main modes: SPEECH and COMFORT_NOISE. The initial mode shall be SPEECH.

The SCR operation on the RX side shall be as follows:

– The RX SCR handler shall enter mode SPEECH, when a frame classified as SPEECH_GOOD is received. Whenever a frame classified as SPEECH_GOOD is received the RX SCR handler shall pass it directly on to the speech decoder.

  If the RX SCR handler is in mode SPEECH, then frames classified as SPEECH_BAD, SPEECH_LOST, or NO_DATA shall be substituted and muted as defined in Appendix I/G.722. Frames classified as NO_DATA shall be handled like SPEECH_LOST frames without valid speech information.

  If the error concealment of RX SCR handler does not support the RX_TYPE=SPEECH_LOST, then frames classified as SPEECH_LOST shall be substituted with RX_TYPE=SPEECH_BAD.

– Frames classified as SID_FIRST, SID_UPDATE or SID_BAD shall bring the RX SCR handler into mode COMFORT_NOISE and shall result in comfort noise generation, as defined in Annex A/G.722.2. SID_BAD frames shall be substituted and muted as defined in Appendix I/G.722.2.

– In mode COMFORT_NOISE the RX SCR handler shall ignore all unusable frames (NO_DATA, SPEECH_BAD); comfort noise generation shall continue, until timeout may apply (Appendix I/G.722.2).

B.4.3 AMR-WB SID information format

When the TX SCR handler is ordered by the network to operate in AMR-WB mode with SCR operation enabled, the SID_UPDATE frame format is according to Annex E/G.722.2. This is the default and only mandatory operating mode of the SCR handler.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series B	Means of expression: definitions, symbols, classification
Series C	General telecommunication statistics
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	**Transmission systems and media, digital systems and networks**
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	TMN and network maintenance: international transmission systems, telephone circuits, telegraphy, facsimile and leased circuits
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks and open system communications
Series Y	Global information infrastructure and Internet protocol aspects
Series Z	Languages and general software aspects for telecommunication systems

Geneva, 2002