



INTERNATIONAL TELECOMMUNICATION UNION

**ITU-T**

TELECOMMUNICATION  
STANDARDIZATION SECTOR  
OF ITU

**H.320**

(03/2004)

SERIES H: AUDIOVISUAL AND MULTIMEDIA SYSTEMS

Infrastructure of audiovisual services – Systems and  
terminal equipment for audiovisual services

---

**Narrow-band visual telephone systems and  
terminal equipment**

ITU-T Recommendation H.320

---

ITU-T H-SERIES RECOMMENDATIONS  
AUDIOVISUAL AND MULTIMEDIA SYSTEMS

CHARACTERISTICS OF VISUAL TELEPHONE SYSTEMS	H.100–H.199
INFRASTRUCTURE OF AUDIOVISUAL SERVICES	
General	H.200–H.219
Transmission multiplexing and synchronization	H.220–H.229
Systems aspects	H.230–H.239
Communication procedures	H.240–H.259
Coding of moving video	H.260–H.279
Related systems aspects	H.280–H.299
<b>Systems and terminal equipment for audiovisual services</b>	<b>H.300–H.349</b>
Directory services architecture for audiovisual and multimedia services	H.350–H.359
Quality of service architecture for audiovisual and multimedia services	H.360–H.369
Supplementary services for multimedia	H.450–H.499
MOBILITY AND COLLABORATION PROCEDURES	
Overview of Mobility and Collaboration, definitions, protocols and procedures	H.500–H.509
Mobility for H-Series multimedia systems and services	H.510–H.519
Mobile multimedia collaboration applications and services	H.520–H.529
Security for mobile multimedia systems and services	H.530–H.539
Security for mobile multimedia collaboration applications and services	H.540–H.549
Mobility interworking procedures	H.550–H.559
Mobile multimedia collaboration inter-working procedures	H.560–H.569
BROADBAND AND TRIPLE-PLAY MULTIMEDIA SERVICES	
Broadband multimedia services over VDSL	H.610–H.619

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

# **ITU-T Recommendation H.320**

## **Narrow-band visual telephone systems and terminal equipment**

### **Summary**

This Recommendation specifies technical requirements for narrow-band visual telephone systems and terminal equipment, typically for videoconferencing and videophone services. It describes a generic system configuration consisting of a number of elements which are specified by respective ITU-T Recommendations, definition of communication modes and terminal types, call control arrangements, terminal aspects and interworking requirements.

This revised version of H.320 introduces a number of enhancements and clarifications to the previous version, primarily the description on the usage of ITU-T Recs G.722.1, H.239, H.241, H.264, and ISO/IEC 14496-3 in H.320 systems.

### **Source**

ITU-T Recommendation H.320 was approved on 15 March 2004 by ITU-T Study Group 16 (2001-2004) under the ITU-T Recommendation A.8 procedure.

## 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.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure e.g. interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

## 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 not 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.

© ITU 2004

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

## CONTENTS

	Page
1 Scope .....	1
2 Definitions .....	1
3 System description.....	2
3.1 Block diagram and identification of elements.....	2
3.2 Signals .....	3
3.3 Bit rate options and infrastructure .....	3
3.4 Call control arrangements.....	6
3.5 Optional enhancements .....	11
4 Terminal requirements.....	11
4.1 Environments.....	11
4.2 Audio and video arrangements.....	12
4.3 Delay compensation in the audio path.....	13
4.4 Control and Indications (C&I).....	14
4.5 Multipoint operation.....	17
5 Intercommunications .....	17
5.1 Intercommunication between different visual telephone terminal types.....	17
5.2 Intercommunication with telephony .....	17
5.3 Intercommunication with other audiovisual terminals .....	18
6 Maintenance.....	19
7 Human factor aspects.....	19
7.1 Physical section .....	19
7.2 Logical section.....	19
Annex A – Optional enhanced video algorithms for H.320 systems.....	20
A.1 Introduction to enhanced video algorithms .....	20
A.2 Use of ITU-T Rec. H.262 in H.320 systems .....	21
A.3 Use of ITU-T Rec. H.263 in H.320 systems .....	23
A.4 Enhanced video algorithm hierarchy .....	24
A.5 Use of ITU-T Rec. H.264 in H.320 systems .....	26



# ITU-T Recommendation H.320

## Narrow-band visual telephone systems and terminal equipment

### 1 Scope

This Recommendation covers the technical requirements for narrow-band visual telephone services defined in H.200/F.720-series Recommendations, where channel rates do not exceed 1920 kbit/s.

NOTE – It is anticipated that this Recommendation will be extended to a number of Recommendations each of which would cover a single videoconferencing or videophone service (narrow-band, broadband, etc.). However, large parts of these Recommendations would have identical wording, while in the points of divergence the actual choices between alternatives have not yet been made; for the time being, therefore, it is convenient to treat all the text in a single Recommendation.

The service requirements for visual telephone services are presented in ITU-T Recs F.720 for videotelephony and F.702 for videoconference; video and audio coding systems and other technical aspects common to audiovisual services are covered in other Recommendations in the H.200/F.700-series.

### 2 Definitions

This Recommendation defines the following terms:

**2.1 bit-rate allocation signal (BAS):** eight bits positioned within the frame structure of ITU-T Rec. H.221 to transmit, e.g., commands, control and indication signals, capabilities.

**2.2 control and indication (C&I):** End-to-end signalling between terminals consisting of control which causes a state change in the receiver and indication which provides for information as to the functioning of the system; see also ITU-T Rec. H.230.

**2.3 data port:** Input/output gate for the user data transmitted within service channel or sub-channels according to ITU-T Rec. H.221.

**2.4 human-machine interface (HMI):** Human-machine interface between user and terminal/system which consists of a physical section (electro-acoustic, electro-optic transducer, keys, etc.) and a logical section dealing with functional operation states.

**2.5 in-band signalling:** Signalling via BAS of the H.221 frame structure.

**2.6 lip synchronization:** Operation to provide the feeling that the speaking motion of the displayed person is synchronized with that person's voice. Alternatively, the minimization of the relative delay between the visual display of a person speaking and the audio of the voice of the person speaking. The objective is to achieve a natural relationship between the visual image and the aural message for the viewer/listener.

**2.7 multipoint control unit (MCU):** A piece of equipment located in a node of the network or in a terminal which receives several channels from access ports and, according to certain criteria, processes audiovisual signals and distributes them to the connected channels.

**2.8 narrow-band:** Bit rates ranging from 64 kbit/s to 1920 kbit/s. This channel capacity may be provided as a single B/H<sub>0</sub>/H<sub>11</sub>/H<sub>12</sub>-channel or multiple B/H<sub>0</sub>-channels in ISDN.

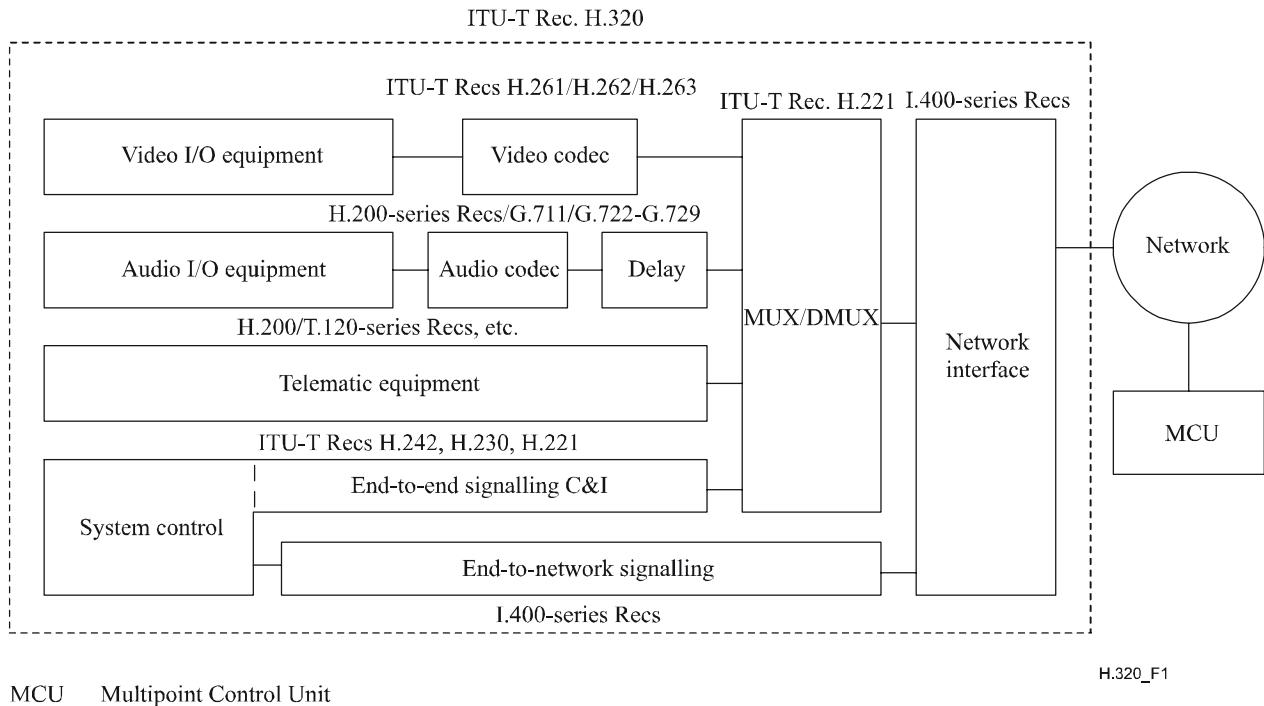
**2.9 out-band signalling:** Signalling via a channel not being part of the B/H<sub>0</sub>/H<sub>11</sub>/H<sub>12</sub>-channel (due to I.400-series Recommendations).

**2.10 visual telephone services:** A group of audiovisual services including videophone defined in ITU-T Rec. F.721 and videoconferencing defined in ITU-T Rec. F.702.

### 3 System description

#### 3.1 Block diagram and identification of elements

A generic visual telephone system is shown in Figure 1. It consists of terminal equipment, network, Multipoint Control Unit (MCU) and other system operation entities.



**Figure 1/H.320 – Visual telephone system**

A configuration of the terminal equipment consisting of several functional units is also shown in Figure 1. Video I/O equipment includes cameras, monitors and video processing units to provide functions such as split-screen scheme. Audio I/O equipment includes microphones, loudspeakers and audio processing units to provide such functions as acoustic echo cancellation (see ITU-T Rec. G.167). Telematic equipment includes visual aids such as electronic blackboard, text conversation facility and still picture transceiver to enhance basic visual telephone communication. The system control unit carries out such functions as network access through end-to-network signalling and end-to-end control to establish common mode of operation and signalling for proper operation of the terminal through end-to-end signalling. The video codec carries out redundancy reduction coding and decoding for video signals, while audio codec does the same thing for audio signals. The delay in the audio path compensates video codec delay to maintain lip synchronization. The mux/dmux unit multiplexes transmitting video, audio, data and control signals into a single bit stream and demultiplexes a received bit stream into constituent multimedia signals. Network interface makes necessary adaptation between the network and the terminal according to the user-network interface requirements defined in the I.400-series Recommendations (see Note).

NOTE – For leased line networks, the network interface is defined in ITU-T Rec. G.703 for bit rates in the range of 64 kbit/s to 2048 kbit/s. An alternative interface is defined in ITU-T Rec. X.21. For  $n \times H_0$  channels, timeslot allocation is given in clause 5/G.704 for the G.703 interface. It is stressed that interworking towards ISDN requires synchronous operation of the leased line network.

### 3.2 Signals

Visual telephone signals are classified into video, audio, data and control as follows:

- Audio signals are continuous traffic and require real-time transmission.  
NOTE – In order to reduce the average bit rate of audio signals, voice activation can be introduced (in which case the audio signals are no longer continuous).
- Video signals are also continuous traffic; the bit rate allocated to video signals should be as high as possible, in order to maximize the quality within the available channel capacity.
- Data signals include still pictures, facsimile and documents, or other facilities such as text conversation; this signal may occur only occasionally as required and may temporarily displace all or part of the audiovisual signal content. It should be noted that data signals are associated only with optional enhancements to the basic visual telephone system; therefore, the opening of a path to carry such signals is preceded by negotiation between the terminals.
- Control signals are some system control signals by definition. The path for the terminal-to-network control signals is provided in the D-channel, while the path for the terminal-to-terminal control signals is provided in BAS or service channel only when necessary by the mechanism defined in ITU-T Rec. H.221.

### 3.3 Bit rate options and infrastructure

#### 3.3.1 Communication modes of visual telephone

Communication modes of visual telephone are defined in Table 1 according to their channel configuration and coding. A particular communication mode of operation is established according to the H.242 procedures.

**Table 1/H.320 – Communication modes of visual telephone**

Visual telephone mode (Suffix n corresponds to audio coding)	Channel rate [kbit/s]	ISDN channel	ISDN interface		Coding		Notes				
			Basic	Primary rate	Audio	Video					
a <sub>n</sub>	64	B	Applicable	n Rec. 0 G.711 1 G.722 2 G.728 3 G.723.1 (Note 5) 4 G.729	Rec. H.261 Rec. H.262 Rec. H.263 (Note 6)	1, 2 3 3, 4 3, 4 3, 4 3 3 3 3 3					
b <sub>n</sub>	128	2B									
c <sub>n</sub>	192	3B									
d <sub>n</sub>	256	4B									
e <sub>n</sub>	320	5B									
f <sub>n</sub>	384	6B									
g <sub>n</sub>	384	H <sub>0</sub>									
h <sub>n</sub>	768	2H <sub>0</sub>									
i <sub>n</sub>	1152	3H <sub>0</sub>									
j <sub>n</sub>	1536	4H <sub>0</sub>									
k <sub>n</sub>	1536	H <sub>11</sub>	Not applicable								
l <sub>n</sub>	1920	5H <sub>0</sub>									
m <sub>n</sub>	1920	H <sub>12</sub>									

**Table 1/H.320 – Communication modes of visual telephone**

Visual telephone mode (Suffix n corresponds to audio coding)	Channel rate [kbit/s]	ISDN channel	ISDN interface		Coding		Notes			
			Basic	Primary rate	Audio	Video				
NOTE 1 – If a visual telephone interworks with a wideband speech terminal, G.722 audio (mode $a_1$ ) may be used instead of G.711 audio (mode $a_0$ ).										
NOTE 2 – If two terminals connect at this rate and run G.711 (mode $a_0$ ) or G.722 (mode $a_1$ ) and both have video capability, H.261, H.262 or H.263 may be used. It should be noted, however, that the video performance is limited due to the very low bit rate available for this purpose.										
NOTE 3 – For multiple channels of $B/H_0$ , all channels are synchronized at the terminal according to 2.7/H.221. For bit rates higher than 64 kbit/s, channel aggregation of multiple B-channels may be used as specified in ITU-T Rec. H.244 or ISO/IEC 13871.										
NOTE 4 – This mode is applicable to the ISDN basic interface if multiple basic accesses are used.										
NOTE 5 – For G.723.1 audio, a channel shall be allocated as specified in ITU-T Rec. H.221. The audio data and their packet formats are described in ITU-T Recs G.723.1 and H.223. Use of bits not required for G.723.1 audio coding is specified in ITU-T Rec. H.221.										
NOTE 6 – The hierarchy of H.261, H.262 and H.263 is defined in Annex A.										

### 3.3.2 Terminal types of visual telephone

The terminal type is categorized according to the type of communication channels with which the terminal can communicate;  $mxB$  (type X with parameter a-f),  $n \times H_0$  (type Y with parameter 1-5; see Note),  $H_{11}/H_{12}$  (type Z with parameter  $\alpha-\beta$ ) or their combinations. Table 2 lists terminal types of visual telephone and communication modes which each terminal type is required to support.

NOTE – Type Y terminals shall have the SM-comp or 6B- $H_0$  compatibility mode defined in ITU-T Rec. H.221 for interworking of evolving networks (see 3.3.2.2).

**Table 2/H.320 – Visual telephone terminal types**

Terminal type		ISDN channel	Modes to be supported		
			Mandatory	Recommended	Optional
X (Note 1)	X <sub>a</sub>	B	$a_0$	–	all other $a_n$
	X <sub>b</sub>	2B	$a_0, b_0$	–	all other $a_n, b_n$
	X <sub>c</sub>	3B	$a_0, b_0, c_0$	–	all other $a_n, b_n, c_n$
	X <sub>d</sub>	4B	$a_0, b_0, c_0, d_0$	–	all other $a_n, b_n, c_n, d_n$
	X <sub>e</sub>	5B	$a_0, b_0, c_0, d_0, e_0$	–	all other $a_n, b_n, c_n, d_n, e_n$
	X <sub>f</sub>	6B	$a_0, b_0, c_0, d_0, e_0, f_0$	–	all other $a_n, b_n, c_n, d_n, e_n, f_n$

**Table 2/H.320 – Visual telephone terminal types**

Terminal type		ISDN channel	Modes to be supported		
			Mandatory	Recommended	Optional
Y (Note 2)	Y <sub>1</sub>	H <sub>0</sub>	g <sub>0</sub>	g <sub>1</sub>	all other g <sub>n</sub>
	Y <sub>2</sub>	2H <sub>0</sub>	g <sub>0</sub> , h <sub>0</sub>	g <sub>1</sub> , h <sub>1</sub>	all other g <sub>n</sub> , h <sub>n</sub>
	Y <sub>3</sub>	3H <sub>0</sub>	g <sub>0</sub> , h <sub>0</sub> , i <sub>0</sub>	g <sub>1</sub> , h <sub>1</sub> , i <sub>1</sub>	all other g <sub>n</sub> , h <sub>n</sub> , i <sub>n</sub>
	Y <sub>4</sub>	4H <sub>0</sub>	g <sub>0</sub> , h <sub>0</sub> , i <sub>0</sub> , j <sub>0</sub>	g <sub>1</sub> , h <sub>1</sub> , i <sub>1</sub> , j <sub>1</sub>	all other g <sub>n</sub> , h <sub>n</sub> , i <sub>n</sub> , j <sub>n</sub>
	Y <sub>5</sub>	5H <sub>0</sub>	g <sub>0</sub> , h <sub>0</sub> , i <sub>0</sub> , j <sub>0</sub> , l <sub>0</sub>	g <sub>1</sub> , h <sub>1</sub> , i <sub>1</sub> , j <sub>1</sub> , l <sub>1</sub>	all other g <sub>n</sub> , h <sub>n</sub> , i <sub>n</sub> , j <sub>n</sub> , l <sub>n</sub>
Z	Z <sub>α</sub>	H <sub>11</sub>	k <sub>0</sub>	k <sub>1</sub>	all other k <sub>n</sub>
	Z <sub>β</sub>	H <sub>12</sub>	m <sub>0</sub>	m <sub>1</sub>	all other m <sub>n</sub>
NOTE 1 – If a visual telephone interworks with a wideband speech terminal, it should support G.722 audio (mode a <sub>1</sub> ) in addition to G.711 audio (mode a <sub>0</sub> ).					
NOTE 2 – Terminal of this type shall conform to 3.3.2.2.					

### 3.3.2.1 Examples

- a) Type X<sub>b</sub> equipped with G.711, G.722 and G.728 audio codecs is a terminal capable of operating at modes a<sub>0</sub>, a<sub>1</sub>, a<sub>2</sub>, b<sub>0</sub>, b<sub>1</sub> and b<sub>2</sub> through B or 2 × B-channel;
- b) Type X<sub>b</sub>Y<sub>1</sub> equipped with G.711, G.722 and G.728 audio codecs is a terminal capable of operating at modes a<sub>0</sub>, a<sub>1</sub>, a<sub>2</sub>, b<sub>0</sub>, b<sub>1</sub>, b<sub>2</sub>, g<sub>0</sub>, g<sub>1</sub> and g<sub>2</sub>, through B, 2 × B- or H<sub>0</sub>-channel;
- c) Type X<sub>f</sub>Y<sub>4</sub>Z<sub>α</sub> equipped with G.711, G.722 and G.728 audio codecs is a terminal capable of operating at modes x<sub>0</sub>, x<sub>1</sub> and x<sub>2</sub> (x = a-k) through (1-6) × B, (1-4) × H<sub>0</sub>- or H<sub>11</sub>-channel.

The type of remote terminal is identified through the transfer rate capability exchange defined in ITU-T Rec. H.242.

### 3.3.2.2 Conditional requirements for single-channel equipment to operate into H.244 Channel Aggregators

In the case where a single channel-equipment, such as a Type Y<sub>1</sub> or Type Z terminal defined here and a terminal which has sub-primary rate channel access capability, operates as an option into a Channel Aggregation Unit (CAU) conforming to H.244 Mode H<sub>2</sub>, the following properties are relevant (see Note):

- 1) The terminal shall be capable of {SM-comp} defined in ITU-T Rec. H.221 – able to vacate bit 8 of the first 16 octets of every timeslot in each frame, in TS2, ..., and to accept an incoming signal of the same structure.
- 2) The terminal shall be able to accept capability sets containing repeated {null} values as described in ITU-T Rec. H.244.
- 3) The terminal should be able to operate at all multiples of 64 kbit/s up to the maximum rate (optional – able to operate with available channel capacity when one or more channels are lost or unavailable).
- 4) The terminal should be able to obey the command [capex] as described in ITU-T Rec. H.244.
- 5) The terminal should be able to recognize and act upon the command [AggIN]\* as described in ITU-T Rec. H.244.

NOTE – The 1993 version, and prior versions of this Recommendation, stipulated {6B-H<sub>0</sub>-comp} rather than {SM-comp} as the capability to be implemented, and there were no provisions for {null}, [capex] and [AggIN]\*. A terminal constructed to one of those versions will interoperate with a Type X<sub>f</sub> terminal via a CAU at 384 kbit/s only, provided that it can accept capssets containing repeated {null} values.

### **3.3.3 Video codec**

As per ITU-T Rec. H.261, and optionally as per H.260-series Recommendations according to the video hierarchy specified in Annex A, and in ITU-T Recs H.241 and H.242.

### **3.3.4 Audio codec**

As per ITU-T Recs G.711, and optionally G.722, G.728, G.723.1, G.729 (see Table 1).

### **3.3.5 Frame structure**

As per ITU-T Rec. H.221.

### **3.3.6 Control and Indication (C&I)**

Identified subset of ITU-T Rec. H.230 is used (see 4.4).

### **3.3.7 Communication procedure**

As per ITU-T Rec. H.242.

## **3.4 Call control arrangements**

To establish intercommunication between various audiovisual terminals, it is necessary to carry out in-band and out-band procedures according to ITU-T Rec. H.242 and other relevant Recommendations.

The different stages of the call are referred according to a point-to-point configuration where terminal X is the calling terminal and Y the called terminal.

### **3.4.1 Establishment of a visual telephone call – Normal procedure**

The provision of the communication is made in the following main steps:

- Phase A: Call set-up, out-band signalling;
- Phase B1: Mode initialization on initial channel;
- Phase CA: Call set-up of additional channel(s), if relevant;
- Phase CB1: Initialization on additional channel(s);
- Phase B2 (or CB2): Establishment of common parameters;
- Phase C: Visual telephone communication;
- Phase D: Termination phase;
- Phase E: Call release.

#### **3.4.1.1 Phase A – Call set-up**

After user initialization, the terminal X performs a call set-up procedure (this procedure is defined in ITU-T Rec. Q.939). As soon as the terminal receives an indication from the network that the connection is established, a bidirectional channel is opened from end to end, and it overlays H.221 framing on the channel.

Following the connection establishment, all the terminals will start to work in Mode 0F (A-law or μ-law) as defined in ITU-T Rec. H.221.

The in-band procedure is activated.

### 3.4.1.2 Phase B1 – Mode initialization

#### 3.4.1.2.1 Phase B1-1

Using the procedures provided in ITU-T Rec. H.242, framed PCM audio is transmitted in both directions (see 3.4.5.1); after frame and multiframe alignment is gained, terminal capabilities are exchanged.

#### 3.4.1.2.2 Phase B1-2 (terminal procedure)

Determination of the appropriate mode to be transmitted: this will normally be the highest common mode (see Table 3 for the case using a B or  $2 \times$  B-channel), but a lower compatible mode could be chosen instead.

In the case that both terminals have announced the capability to work on additional channel(s), terminal X initiates the request for additional call set-up.

Visual telephone terminals using multiple connections need to know what number(s) to dial for additional connection(s). In many cases this information is stored, or can be deduced from the number dialled for the initial channel. Where the information is not available or deducible, the terminal can acquire the missing number(s) by the procedure defined in 18/H.242.

NOTE – If the user at either terminal does not wish the call to proceed to two or more channels, even though the terminal has this capability, the user must set the terminal such that only single-channel capability is declared in phase B1-1. In this case, we distinguish the active capability, wished by the users, from the inherent capability of the terminal. See clause 5/H.242.

**Table 3/H.320 – Common mode (default) for communication between different types of visual telephones using a B or  $2 \times$  B-channel**

<b>X<sub>a</sub> (G.711, G.728)</b>	<b>X<sub>b</sub> (G.711)</b>	<b>X<sub>b</sub> (G.711, G.722)</b>	<b>X<sub>b</sub> (G.711, G.728)</b>	<b>X<sub>b</sub> (G.711, G.722, G.728)</b>	<b>Terminal type (equipped audio codecs)</b>
a <sub>2</sub>	a <sub>0</sub>	a <sub>0</sub>	a <sub>2</sub>	a <sub>2</sub>	X <sub>a</sub> (G.711, G.728)
	b <sub>0</sub>	b <sub>0</sub>	b <sub>0</sub>	b <sub>0</sub>	X <sub>b</sub> (G.711)
		b <sub>1</sub>	b <sub>1</sub>	b <sub>1</sub>	X <sub>b</sub> (G.711, G.722)
			b <sub>2</sub>	b <sub>2</sub>	X <sub>b</sub> (G.711, G.728)
				b <sub>1</sub> /b <sub>2</sub>	X <sub>b</sub> (G.711, G.722, G.728)

NOTE 1 – "b<sub>1</sub>/b<sub>2</sub>" indicates that either b<sub>1</sub> or b<sub>2</sub> can be used. b<sub>1</sub> gives better audio, while b<sub>2</sub> gives better video.

NOTE 2 – The communication modes indicated in this table include the possibility to use the H.261 CIF format as well as the H.261 QCIF format. Optionally, other modes defined in the H.260-series Recommendations may be used. The choice of the operating format is left to the terminal procedure according to the exchanged capabilities: symmetry in two directions is not essential.

Each terminal may use a minimum decodable picture interval in its sending direction which makes best use of the capability indicated by the other terminal.

This table does not include interworking situations between visual telephones and telephone terminals. If visual telephone terminals are connected to telephones, mode a<sub>0</sub> is used for the communication.

### **3.4.1.2.3 Phase B1-3 (mode switching)**

If the terminal has {G.728}, {G.723.1}, or {G.729} capability (or some combination of these capabilities) and is connected to a 64/56 kbit/s connection, and if the capset received includes {1B}, {G.728}, {G.723.1}, or {G.729} and a video capability for H.261, H.262, or H.263, identifying the remote endpoint as Type X<sub>a</sub>, it shall switch to one of Modes a<sub>2</sub>, a<sub>3</sub>, or a<sub>4</sub>. Likewise, if the terminal is itself Type X<sub>a</sub> and the received capset includes {G.728}, {G.723.1}, or {G.729} (or some combination of these capabilities) and a video capability for H.261, H.262, or H.263, it shall switch to one of Modes a<sub>2</sub>, a<sub>3</sub>, or a<sub>4</sub>.

In all other cases a higher capacity than 64/56 kbit/s is available, and the terminal may switch to b<sub>0</sub> or a higher mode in Table 1, according to the application.

If the terminal is of Type X<sub>b</sub>, X<sub>c</sub> or higher and the received capset includes {2B} or higher, it is a matter for local choice whether to mode-switch video on and/or change to another audio coding mode while waiting for additional connections to be set up.

The switching is made using the procedure of ITU-T Rec. H.242. If the terminals have not both adopted the common mode, an asymmetric communication may result (this is not precluded) – see 3.4.1.5.

### **3.4.1.3 Phase CA – Call set-up of the additional channel(s)**

Following phases B1-3 and B2 if relevant, the communication phase C proceeds on that channel. If additional channels have been requested, these also go through phase A (hence the nomenclature "Phase CA"), exactly as in phase A above, and additional call set-ups are performed by the terminals. On each of the established channels, H.221 framing is overlaid (see Note).

NOTE – During phase CA, an intermediate audiovisual mode could be offered on the initial channel used for initialization, until full completion of initialization phase.

### **3.4.1.4 Phase CB1 – Mode initialization on additional channel(s)**

#### **3.4.1.4.1 Phase CB1-11**

Using the procedure provided in ITU-T Rec. H.242, frame and multiframe alignments are gained.

#### **3.4.1.4.2 Phase CB1-12**

Synchronization of the channels is achieved.

#### **3.4.1.4.3 Phase CB1-2 (terminal procedure)**

Determination of the appropriate mode to be transmitted. This phase will normally be skipped, because the appropriate mode of operation is determined in Phase B1-2.

#### **3.4.1.4.4 Phase CB1-3 (mode switching)**

Both terminals switch to the mode they have identified in phase B1-2 using the procedure of ITU-T Rec. H.242.

It may be noted that G.711 modes should be avoided after initialization if ITU-T Rec. G.728 or G.722 is available, since both these can leave more capacity for video and ITU-T Rec. G.722 provides better quality.

Total symmetry is not required: that is, it is not essential that the audio and video transmissions be identical in the two directions, and 8.2/H.242 warns against implementations that automatically select the outgoing mode to be the same as the incoming, listing better methods of mode control. Visual telephone applications require symmetry of transfer rate, but choice of transmitted audio and video coding settings are made by the user, or preset in the terminal.

### **3.4.1.5 Phase B2 (or CB2) – Establishment of common parameters**

This phase establishes common operational parameters specific to visual telephone (e.g., encryption) after phase B1 process is finished. Capabilities or requirements of the receiving side are first indicated, then the sending side decides operational parameters and controls the receiving side. BAS codes for this purpose are defined in ITU-T Rec. H.221.

Sometimes users may wish to indicate preference of receiving modes (e.g., choice between mode  $b_2$  with better audio quality and mode  $b_3$  with better video quality as in Table 3). This can be achieved by using the "mode-preference" indications defined in 9.5/H.242. It is recommended that, in general, visual telephones be equipped to send and respond to these codes.

### **3.4.1.6 Phase C – Visual telephone communication**

In the case where more than one channel is used, there will be intermediate phases CA, CB1, CB2 as described in this clause. Likewise, if additional channels are dropped during the call, there will be intermediate phases CD, CE as described in 3.4.4. The provisions of this clause apply to any channel, initial or additional, for which phases B1 and B2 have been completed and phase D not yet started.

#### **3.4.1.6.1 Mode switching**

According to action by either user (for example, starting a facsimile machine), a different mode from the highest common mode may become more appropriate. Switching to this mode is made according to the procedure of ITU-T Rec. H.242.

#### **3.4.1.6.2 Capability change**

The user may change the capability of his terminal during the call (for example, by connecting or switching-on auxiliary telematic equipment); the terminal must initiate the capability exchange procedure defined in ITU-T Rec. H.242.

### **3.4.1.7 Phase D – Termination phase**

#### **3.4.1.7.1 Phase D1 (terminal procedure)**

When one of the users hangs up, the terminal invokes phase D2 directly.

#### **3.4.1.7.2 Phase D2 (mode switching)**

Mode 0F is forced according to ITU-T Rec. H.242 (or taking into account the result of phase D1 if different; for further study).

In case the call is disconnected for some reason, it may take a long time if it keeps awaiting reaction from the remote terminal, and reception or initiation of the next call may be blocked. ISDN disconnection should be given priority over the H.242 call termination procedure. A time-out of two seconds is suggested.

### **3.4.1.8 Phase E – Call termination (release)**

The terminal which has initiated the hang up sends messages over the D-channel with respect to all channels and idles all of them (that means no more information is sent over).

At the other terminal, the actual disconnection occurs at reception of the other disconnect message(s).

### **3.4.2 Exceptional procedures for phases A and B**

In case of unsuccessful outcome during phases A and B (due to many causes), exceptional procedures may be provided in order to ensure a suitable service. The matter is for further study.

### **3.4.3 Exceptional procedures during phase C**

During the actual exchange of audiovisual data, problems may occur in some channels. Fallback procedures managed by the terminal are activated. The description of the procedures and the appropriate indications are for further study.

### **3.4.4 Addition and dropping of channels during a visual telephone call**

#### **3.4.4.1 Addition**

According to action by a user (for example the activation of auxiliary equipment), one or more additional channels are requested. The procedure follows those described for phases CA and CB1.

#### **3.4.4.2 Dropping**

Two phases are envisaged.

##### **3.4.4.2.1 Phase CD1**

The common mode, appropriate to the channel(s) which remains, is selected.

##### **3.4.4.2.2 Phase CD2**

The mode switching procedure of ITU-T Rec. H.242 is applied to invoke the mode identified in phase CD1; the remaining channel is the channel used for initialization (see phase A). It supports an appropriate fallback mode. The matter is for further study.

### **3.4.5 Transmission and presentation of audiovisual information at the start of a visual telephone call**

#### **3.4.5.1 Audio**

Transmitted audio shall be according to 9.1.1.2/H.242. For videotelephony, audio shall not be muted at the transmitting side.

As soon as the initial channel is established and the received audio coding law (G.711 A-law or  $\mu$ -law) is known according to the procedure defined in 9.1.1.3/H.242, the audio signal should be presented to the user so that audio conversation can promptly be started as in the ordinary telephone communication.

#### **3.4.5.2 Video**

According to the chosen terminal procedures, pictures may or may not be visible to both users as soon as initialization is complete. In the case that either phase B1-3 or phase CB1-3 has activated a common mode, including video, mutual visibility of the users is possible.

The following items collect alternative procedures which can be used to suspend picture display until user intervention (by mutual agreement or otherwise) causes pictures to be displayed.

- 1) *No video transmitted* – In phase B1-2 and (if relevant) phase CB1-2, the mode selected includes video OFF. During phase C either user may unilaterally switch to video ON (H.261 or other H.260-series); alternatively, the terminal may send the C&I BAS code VIR (Video Indicate Ready-to-activate), but not switch to video ON until video or VIR is received from the other terminal. While the incoming video OFF state remains, the visual telephone screen should display a symbol or message indicating this (i.e. there is no fault).

As already noted in 3.4.1, phase B1-2, the request for additional channel may, according to terminal procedure, be delayed while video OFF is maintained; user action to activate video would then result in procedure phases CA1, CB1 (CB2 if required).

- 2) *Video pattern transmitted* – An electronically generated or other pattern is transmitted instead of the signal from a normal camera. The C&I BAS code VIS (Video Indicate Suppressed) is used to indicate the situation to the remote party.

- 3) *Video transmitted but not displayed* – Terminal procedures simply involve local action to display not the incoming signal but an explanatory symbol or message. User action would cause the incoming signal to be displayed, but if this should depend on mutual action by both users, then a new C&I BAS code VRD (Video Ready-to-Display) must be defined. This point is for further study.

### 3.5 Optional enhancements

#### 3.5.1 Data ports

Data ports as physical I/O ports of the terminal for telematic and other equipment are activated/deactivated by BAS commands. Depending on the transmission capability of a connection, e.g., multiples of B/H<sub>0</sub> channels, etc., various bit rates are available at these ports. Allocation of bit streams to the port(s) is performed by in-band signalling. Data conveyed at the port(s) is transparent, data rates being listed in Annex A/H.221.

For the data enhancement of the conformant visual telephone terminals, the following applies for increased level of interworking:

- 1) Terminals having at least one "audiographic" application such as T.81 still picture exchange, application sharing, on-screen annotation, text conversation etc. shall make use of the standardized applications provided by the T.120-series (if such applications have been recommended by the ITU-T) making use of the MLP/H-MLP channels. A terminal that provides far-end camera control using ITU-T Recs H.281 and H.224 operating in both the LSD and the MLP channels is not required to also support a T.120 far end camera control protocol. Similarly, a terminal that provides text conversation using ITU-T Recs T.140 and H.224 operating in both the LSD and the MLP channels is not required to also support a T.120 text conversation facility.
- 2) Preferred MLP channel rates are 6.4 kbit/s (5.6 kbit/s if ECS is in use), 14.4 kbit/s (13.6 kbit/s if ECS is in use), 32 kbit/s and 40 kbit/s with 6.4 kbit/s (5.6 kbit/s if ECS is in use) as default.
- 3) If a terminal has an application which uses LSD to obtain critical performance, it should also be able to transmit the requisite data in the MLP channel when in multipoint calls via an MCU.

#### 3.5.2 Encryption

Encryption may be applied on audio and video signals multiplexed; the encryption shall be according to ITU-T Rec. H.233 and, where a key-management system is required, one of the schemes in ITU-T Rec. H.234 should be used. Switching-on and off the encryption process has to be signalled between the terminals (or terminal and MCU, respectively) via in-band signalling.

#### 3.5.3 Restricted networks

Some networks are restricted in transfer characteristics. For communications among terminals accommodated in restricted networks and those accommodated in non-restricted networks, both in point-to-point and multipoint cases, optional procedures are defined in clause 13/H.242.

## 4 Terminal requirements

### 4.1 Environments

Under study.

## **4.2 Audio and video arrangements**

### **4.2.1 Audio arrangements**

A terminal can have one or more of three different arrangements:

- handset function;
- handsfree function for a small group of users (up to three users);
- handsfree function for more than three users (conference terminal).

The audio characteristics are defined for each of these functions. Furthermore, the bandwidth of the transmitted speech is taken into consideration.

The principles used are identical with those for telephony terminals. That is, the sensitivity for handset function and handsfree function designed for personal use, or for a small group of users, is specified in loudness ratings, and the sensitivity for conference terminals is specified as output levels.

#### **4.2.1.1 Test principles**

##### **4.2.1.1.1 Handset function**

The sensitivity measurement of a terminal when a handset is used shall be based on the principles described in ITU-T Rec. P.64. The loudness rating shall be calculated as described in ITU-T Rec. P.79.

##### **4.2.1.1.2 Handsfree function for a small group of users**

The sensitivity measurement of the handsfree function of a terminal designed for a small group of users shall be based on the principles described in ITU-T Rec. P.34. The applied test signal level at the digital input when measuring receive sensitivity shall be –30 dBm0.

The user position for a visual telephone terminal depends on the design of the terminal. The real user position as recommended by the supplier might be different compared with the position used for measurements. A correction factor shall be used. The correction factor is:

$$F(dB) = 20 \log_{10} \left\{ \frac{D_s}{D_0} \right\}$$

where

$D_s$  is the distance between the recommended user position and the terminal and

$D_0$  is the reference distance of 50 cm.

The loudness rating shall be calculated as described in ITU-T Rec. P.79.

##### **4.2.1.1.3 Handsfree function for a conference terminal**

The principles described in ITU-T Rec. P.30 shall be used.

#### **4.2.1.2 Sensitivity**

##### **4.2.1.2.1 General**

For handset terminals and handsfree terminals designed for a small group of users, the sensitivity shall be specified as loudness ratings: Send Loudness Rating (SLR) and Receive Loudness Rating (RLR). The definitions of SLR and RLR are found in ITU-T Rec. P.10.

For conference terminals, the sensitivity shall be specified in terms of input and output levels.

#### **4.2.1.2.2 Receive volume control**

For handsfree and loudspeaking terminals, a volume control shall be provided.

Where a manual receive volume control is provided, the minimum control range shall be to  $-15$  dB from the test position.

Where an automatic receive volume control is provided, the RLR value obtained with a line level of  $-15$  dBm<sub>0</sub> shall not exceed that RLR value which is obtained with a line level of  $-30$  dBm<sub>0</sub> by more than  $15$  dB.

#### **4.2.1.2.3 Handset function**

The requirements of Table 4 shall be met.

**Table 4/H.320 – Sensitivity of the handset function**

	<b>3.1 kHz bandwidth (Note)</b>	<b>7 kHz bandwidth</b>
SLR	8	8
RLR	2	7

NOTE – 3.1 kHz bandwidth includes both G.711 and G.728 coding.

The manufacturing tolerances are  $\pm 3$  dB.

#### **4.2.1.2.4 Handsfree function**

The requirements of Table 5 shall be met.

**Table 5/H.320 – Sensitivity of the handsfree function**

	<b>3.1 kHz bandwidth (Note)</b>	<b>7 kHz bandwidth</b>
SLR	13 – F	13 – F
RLR	$-7$ – F	$-5$ – F

NOTE – 3.1 kHz bandwidth includes both G.711 and G.728 coding.

The receive RLR requirement shall be met when the receive volume control is in its maximum position. The manufacturing tolerances are  $\pm 4$  dB.

#### **4.2.1.2.5 Conference terminals**

The procedures and values specified in ITU-T Rec. P.30 shall be used.

### **4.2.2 Video arrangements**

Under study.

## **4.3 Delay compensation in the audio path**

The H.261 and other H.260-series video codecs require some processing delay, while the H.200-series/G.711 and G.722-G.729 audio codecs involve much less delay. Hence, unless the audio is further delayed, the sound will be presented significantly in advance of the lip movements of the speaker. If lip synchronization is to be maintained, the video processing delay must be compensated in the audio path. Since video coder and decoder delays may vary according to implementation, delay compensation must be carried out individually at the coder and decoder. A reference measurement method of video coder and decoder delays is defined in ITU-T Rec. H.261.

The insertion of lip-synch delay is not mandatory: some users may prefer the immediate speech for its better interactivity, while others prefer lip synchronization, accepting the delay it causes. If one user selects delay insertion, the terminal should send the code ACE (see ITU-T Rec. H.230) to request that the remote end-point do likewise; if zero-delay is selected, ACZ should be sent.

When delay is inserted at the transmitting end, care should be taken not to insert a disturbing noise into the encoded audio path, since the receiving end will not be able to mute this out, not knowing exactly when it will occur. When delay is inserted at the receiving end, similar care should be taken to avoid disagreeable noise presentation.

If delay is inserted in the transmission direction, it should also be inserted at the same time in the receiving direction. The insertion could be made:

- at the very start of Phase B (note that if the remote end is not a videophone, the delay is then inappropriate);
- when incoming frame structure is detected (the remote end is then more likely to be a videophone, but could also be an audiographic terminal, for which the delay is then inappropriate);
- when switching video ON [note that this choice may disturb (twice over) the user's initial greetings, and so should be avoided];
- later in the call, or not at all.

#### 4.4 Control and Indications (C&I)

H.221 frame synchronous C&Is are chosen from the general audiovisual set contained in ITU-T Rec. H.230. For visual telephone systems, those signals in Table 6 are mandatory, where their source, sink, synchronization with picture, transmission channel and codewords are indicated. Additionally, support of C&I signals listed in Table 7 is highly desirable for multipoint communication.

All visual telephone terminals have a video source providing a picture of participants, and some terminals may have additional video sources; the participant-picture source is designated #1, having the associated symbol VIA. When incoming video is ON (BAS command (010) [1, 2, 8 or 9]) and VIA, VIA2, VIA3 have not been transmitted, source #1 is assumed.

Some supplementary services provided by some conference bridges, servers, etc. may require terminals to send out the symbols on their dialling keyboards. It is highly desirable that all visual terminals be able to translate these symbols into BAS codes representing the alphabet as provided for in ITU-T Rec. H.230. This functionality is essential for displaying texts input by the user in such applications as videophone for handicapped people.

**Table 6/H.320 – C&I signals mandatory for visual telephone**

	C&I signal	C/I	Source	Sink	Sync. with picture	Transmission channel	Codeword definition
Video	Picture format	I	Decoder	Coder	No	BAS	H.221
	Picture format	C	Coder	Decoder	Yes	Embedded in video	H.261/H.26x <sup>c)</sup>
	Minimum decodable picture interval	I	Decoder	Coder	No	BAS	H.221
	Freeze picture request control, VCF	C	Coder or MCU	Decoder	No	BAS	H.221
	Fast update request control, VCU	C	Decoder or MCU	Coder	No	BAS	H.221
	Freeze picture release control	C	Coder	Decoder	Yes	Embedded in video	H.261/H.26x <sup>c)</sup>
MCU	Multipoint command conference, MCC and cancel-MCC	C	MCU	Terminal	No	BAS	H.230
	Multipoint command symmetrical data transmission, MCS	C	MCU	Terminal	No	BAS	H.230
	Multipoint command negating MCS, MCN	C	MCU	Terminal	No	BAS	H.230
	Multipoint Mode Symmetrize, MMS and cancel-MMS <sup>a)</sup>	C	MCU <sup>b)</sup>	Terminal	No	BAS	H.230
Maintenance	Video loop request control, LCV	C	Terminal	Terminal	No	BAS	H.221
	Digital loop request control, LCD	C	Terminal	Terminal	No	BAS	H.221
	Loop off request, LCO	C	Terminal	Terminal	No	BAS	H.221
Conference	Split screen indication (Note)	I	Sending terminal	Receiving terminal	Yes	Embedded in video	H.261/H.262/ H.263

**Table 6/H.320 – C&I signals mandatory for visual telephone**

	C&I signal	C/I	Source	Sink	Sync. with picture	Transmission channel	Codeword definition
Terminal	Document camera indication (Note)	I	Sending terminal	Receiving terminal	Yes	Embedded in video	H.261/H.262/ H.263
	Audio active/muted indication, AIA/AIM	I	Sending terminal	Receiving terminal	No	BAS	H.230
	Video active indication, VIA	I	Sending terminal	Receiving terminal	No	BAS	H.230
	Video suppressed indication, VIS	I	Sending terminal	Receiving terminal	No	BAS	H.230
<p><sup>a)</sup> MMS and cancel-MMS are mandatory only when H.263 or H.262 and H.263 are supported in addition to H.261.</p> <p><sup>b)</sup> MMS (Multipoint Mode Symmetrize – see clause 4/H.243) and cancel-MMS are commands issued by an MCU to enforce mode symmetry.</p> <p><sup>c)</sup> "H.26x" means any video coding Recommendation of the H.260-series</p> <p>NOTE – The procedures to use these indication signals are under study. However, H.261, H.262 and H.263 decoders shall be able to decode and ignore these indication signals.</p>							

**Table 7/H.320 – Optional C&I signals which should be supported**

	C&I signal	C/I	Source	Sink	Sync. with picture	Transmission channel	Codeword definition
MCU	Multipoint indication zero-communication, MIZ	I	MCU	Terminal	No	BAS	H.230
	Multipoint indication secondary-status, MIS	I	MCU	Terminal	No	BAS	H.230
Video	Video command reject, VCR		C	MCU	Terminal or MCU	No	BAS
Audio	Audio command equalize, ACE		C	Terminal	Terminal	No	BAS
	Audio command zero-delay, ACZ		C	Terminal	Terminal	No	BAS

**Table 7/H.320 – Optional C&I signals which should be supported**

	C&I signal	C/I	Source	Sink	Sync. with picture	Transmission channel	Codeword definition
Mode preference	Mode preference indicators (A-law 0F, µ-law 0F, G.722-m2, G.722-m3, G.728, G.723.1, G.729, H.261/QCIF, H.261/CIF, H.262S SIF, H.262S 2SIF, H.262S 4SIF, H.262M SIF, H.262M 2SIF, H.262M 4SIF, H.263 SQCIF, H.263 QCIF, H.263 CIF, H.263 4CIF, H.263 16CIF, CSFMT, CPAR, CPCF, VSTRD, VSTRDEL, SCLPREF, GHOP, Cancel-GHOP)	I	Terminal	Terminal	No	BAS	H.230

#### 4.5 Multipoint operation

It is highly desirable that terminals which do not have any auxiliary data equipment should nevertheless be able to open LSD and/or HSD as well as MLP and/or H-MLP channels so that they could participate in multipoint communications without losing video. All conforming terminals should be able to declare the "Nil\_Data" capability defined in ITU-T Rec. H.221 and use the procedure defined in 12.5/H.242 for this purpose.

Other aspects of multipoint operation of the terminal are under study.

### 5 Intercommunications

The mechanisms for intercommunication with other services are described in the H.200/H.240-series and G.725, G.233 Recommendations.

#### 5.1 Intercommunication between different visual telephone terminal types

A common mode of operation is determined as described in 3.4.1. D-channel signalling should include new LLC and HLC which are appropriate for audiovisual services, but this clause is for further study.

#### 5.2 Intercommunication with telephony

NOTE – Description of this clause is for communications using a B-channel.

### **5.2.1 Intercommunication with ISDN telephones**

A call from a visual telephone to an ISDN telephone is first placed as an audiovisual call, but the ISDN telephone returns incompatible destination or the network returns recovery on timer expiry in case of no responses from the called side, then the visual telephone may switch to a speech or 7 kHz audio bearer service call.

A call from ISDN telephone to a visual telephone is accepted by the visual telephone because every audiovisual terminal is equipped with this telephone capability as a minimum function.

For both of the above cases, the operational mode of communication is G.711 speech or G.722 audio.

### **5.2.2 Intercommunication with PSTN telephones**

A call from visual telephone to a PSTN telephone may be initiated as an audiovisual call, but the network returns no route to destination, then the visual telephone may switch to a speech or 3.1 kHz audio bearer service call. The operational mode of communication is G.711 audio coding.

A call from a PSTN telephone is routed into the ISDN as a 3.1 kHz audio call which can be responded by the visual telephone for the same reason as described in 5.2.1. The operational mode of communication is 3.1 kHz audio.

## **5.3 Intercommunication with other audiovisual terminals**

A common mode of operation is determined according to the H.200-Series Recommendations.

### **5.3.1 Intercommunication with H.324 terminals on GSTN**

This intercommunication can be achieved using an interworking adapter or using dual-mode (N-ISDN and GSTN) terminals on the N-ISDN. An H.324/H.320 interworking adapter is located at the interface between ISDN and GSTN signals. The operation of the interworking adapter is described in 8.2/H.324. Additional optional functionalities of H.320 terminals to support this intercommunication are under study.

### **5.3.2 Intercommunication with H.322 terminals on Guaranteed QoS LANs**

This intercommunication can be achieved using the H.322 gateway which is described in clause 5/H.322.

### **5.3.3 Intercommunication with H.323 terminals on Non-guaranteed QoS LANs**

This intercommunication can be achieved using the H.320-H.323 gateway which is described in 6.3/H.323 and 9.2/H.323. Additional optional functionalities of H.320 terminals to support this intercommunication are under study.

### **5.3.4 Intercommunication with H.310/H.321 terminals on ATM networks**

Intercommunication between H.310 and H.320 terminals is achieved using the H.320/H.321 interoperation mode defined in ITU-T Rec. H.310. The intercommunication with H.320 terminal is described in 12.1/H.310 and 12.2/H.310. It is through the I.580 interworking function for H.310 RAST-1 terminals, RAST-1&5 terminals working in AAL1 and H.321 terminals, or through the gateway located in the customer premises ATM network for RAST-5 terminals and RAST-1&5 terminals working in AAL5.

## **6 Maintenance**

Some loop-back functions are envisaged to allow verification of the functional aspects of the terminal in order to ensure correct operation of the system and satisfactory quality of the service to the remote party. The following loop-back functions (see Figure 2) are envisaged:

- a) *Loop at terminal-network interface (towards network)*

Upon receiving the digital loop-back BAS, loop-back is activated at the digital interface of the terminal toward the network side. In case of a multiple B/H<sub>0</sub> channel arrangement, loop-back is activated in each connection.

- b) *Loop at terminal-network interface (towards terminal)*

The procedure is for further study.

- c) *Loop at analogue I/O interface*

Upon receiving the video loop-back or audio loop-back BAS, loop-back is activated at the analogue interface of the video/audio codec towards the video/audio codec.

The opportunity of having a self-checking procedure at terminal stage is for further study.

## **7 Human factor aspects**

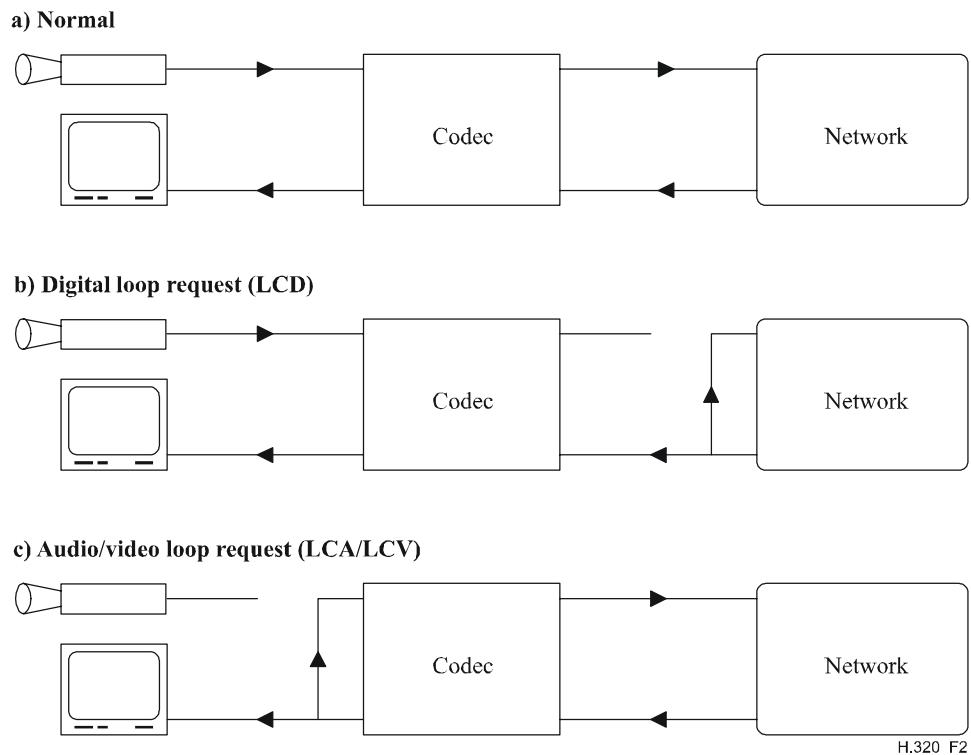
To achieve error-free and uncomplicated utilization of terminal equipment and service from the users' standpoint, human factor related aspects have to be studied and recommended. These aspects deal with the flow of information between user and terminal/network. This information can be divided into a physical section and a logical section of the HMI.

### **7.1 Physical section**

- Figures and properties of transducers (camera, microphone, etc.).
- Signals particularly related to the service, keys, pictograms.

### **7.2 Logical section**

- Procedures, e.g. for call establishment/release, during communication phase.
- Consistency between the MMIs of visual telephone and terminals of other teleservices.



**Figure 2/H.320 – Loop back**

## Annex A

### Optional enhanced video algorithms for H.320 systems

#### A.1 Introduction to enhanced video algorithms

Although the video coding algorithm specified in ITU-T Rec. H.261 has been the recommended video codec for H.320 systems, recognition of the improved flexibility and quality achieved by ITU-T Recs H.262, H.263, and H.264 has made it necessary to introduce these codecs as optional modes for H.320 systems. Among the advantages for H.262 are:

- ability to handle interlaced video;
- ability to utilize 625/50 and 525/60 formats without standards conversion.

Advantages noted for H.263 are:

- coding performance is improved over H.261 with the baseline H.263 algorithm; significant improvement when coding options are utilized;
- multiple image formats.

Advantages noted for H.264 are:

- substantially increased coding efficiency;
- enhanced robustness to network environments.

Detailed information on these optional video algorithms and their implementation in H.320 systems follows in the next clauses.

## A.2 Use of ITU-T Rec. H.262 in H.320 systems

### A.2.1 Image formats

ITU-T Rec. H.262 is a toolkit of compression techniques in which the demands of various applications can be met by selecting the proper tools. This is facilitated by the concept of Profiles and Levels. A Profile is a defined subset of the entire bitstream syntax. A Level is a defined set of constraints imposed on parameters in the bitstream.

The most appropriate Profile and Level for video conferencing within this Recommendation is the Simple Profile at Main Level (SPML). The Simple Profile prohibits the use of B-frames (bidirectionally predicted) specifically because of the added delay which is inherent in their use. The Main Level of the Simple Profile (the only level defined for that profile) indicates that the maximum image sampling is 720 samples/line, 576 lines/frame, and 30 frames/second.

Though the SPML offers the ability to minimize delay, there are situations in which additional delay is acceptable. In fact, an H.262 encoder which uses one or two B-frames between P-frames can achieve delay values which are comparable to those of H.261 encoders with Minimum Picture Intervals (MPIs) of 2 or 3, respectively. For this reason, the Main Profile, Main Level (MPML) is also allowed within this Recommendation. No Profiles and Levels other than SPML and MPML are allowed for use in this Recommendation, i.e., other Profiles and Levels are forbidden.

Since only the maximum image size is specified, SPML and MPML are capable of supporting a large number of image sizes. For the sake of interoperability, the number of possible formats is restricted. The following image formats are allowed in H.320 systems:

SIF	352 × 288 × 25 fps
	352 × 240 × 29.97 fps
	352 × 288 × 29.97 fps
2SIF	352 × 576 × 25 fps
	352 × 480 × 29.97 fps
4SIF	704 × 576 × 25 fps
	704 × 480 × 29.97 fps

The 2SIF and 4SIF formats are intended to be interlaced picture formats with field rates twice the stated frame rates or progressive formats at the stated frame rates. Only the 4:2:0 chroma format is allowed for each of these image formats.

### A.2.2 Forward error correction

H.262 shall always be transmitted using BCH error correction coding in H.320 systems. (For details on BCH error correction, refer to 5.4/H.261.) When H.262 is used in H.320, BCH error correction is mandatory for both encoders and decoders.

### A.2.3 Syntax and semantics of video frame synchronous C&I

In H.262 video streams, the functionalities of the H.261 PTYPE field (i.e. video frame synchronous C&I) is provided in ITU-T\_extension() which uses the ITU-T Extension ID defined in Table 6-2/H.262 [the generic usage of ITU-T\_extension() is provided in A.2.4]. The syntax of ITU-T\_extension() for H.320 video frame synchronous C&I is shown in Table A.1.

**Table A.1/H.320 – Syntax of ITU-T\_extension() for H.320 video frame synchronous C&I**

ITU-T_extension()	No. of bits	Mnemonic
extension_start_code_identifier	4	uimsbf
ITU-T_application_identifier	8	uimsbf
Split_screen_indicator	1	uimsbf
document_camera_indicator	1	uimsbf
freeze_picture_release	1	uimsbf
Next_start_code()		
}		

NOTE – In ITU-T Rec. H.262, image formats are not indicated as video frame synchronous information as in ITU-T Rec. H.261, but are indicated in sequence\_header() and sequence\_header\_extension().

The semantics of each field in Table A.1 is as follows:

extension\_start\_code\_identifier – As shown in Table 6-2/H.262, the value '1100' in binary is used for ITU-T\_extension().

ITU-T\_application\_identifier – A unique 8-bit value for this field is assigned to each ITU-T application for which the information following this field is relevant. The value 0x01 in hexadecimal is assigned to H.320 video frame synchronous C&I.

split\_screen\_indicator, document\_camera\_indicator, and freeze\_picture\_release – The semantics of these bits is identical with the corresponding bits in PTYPE of ITU-T Rec. H.261.

ITU-T\_extension() for H.320 terminals is only transmitted when the information must be changed from its current value. Still image transmission using the format specified in Annex D/H.261 is not an allowable mode of ITU-T Rec. H.262.

#### A.2.4 Generic usage of ITU-T\_extension()

As defined in Table 6-2/H.262, a four bit value '1100' in binary is reserved for H.262 extension\_start\_code\_identifier as ITU-T Extension ID. The ITU-T Extension ID is used in ITU-T\_extension(), which is an H.262 extension field for ITU-T applications.

As defined in 6.2.2.2.1/H.262, ITU-T\_extension() is an extension to picture\_coding\_extension() of ITU-T Rec. H.262. The generic syntax of ITU-T\_extension() is shown in Table A.2.

**Table A.2/H.320 – Generic syntax of ITU-T\_extension()**

ITU-T_extension()	No. of bits	Mnemonic
extension_start_code_identifier	4	uimsbf
ITU-T_application_identifier	8	uimsbf
ITU-T_application_data()		
next_start_code()		
}		

The semantics of each field in Table A.2 is as follows:

extension\_start\_code\_identifier – As shown in Table 6-2/H.262, the value '1100' in binary is coded in this field as ITU-T Extension ID.

ITU-T\_application\_identifier – This field specifies the ITU-T application for which the information in the following ITU-T\_application\_data() field is relevant. A unique and exclusive 8-bit value for this field is assigned to each ITU-T application that uses ITU-T\_extension().

`ITU-T_application_data()` – This field includes the information used by the ITU-T application specified by `ITU-T_application_identifier`. The syntax and semantics of this field are defined in the specifications of each ITU-T application.

`next_start_code()` – The definition of this function is provided in 5.2.3/H.262.

The following items need to be defined in the specifications of each ITU-T application which uses `ITU-T_extension()`:

- `ITU-T Application ID`  
A unique and exclusive 8-bit value for `ITU-T_application_identifier` should be assigned by the application.
- `Syntax and semantics for ITU-T_application_data()`  
The syntax and semantics of `ITU-T_application_data()` is defined in the specifications of each ITU-T application with a unique value for `ITU-T_application_identifier`. In designing this syntax, the occurrence of start code emulation should be carefully avoided.

### A.3 Use of ITU-T Rec. H.263 in H.320 systems

ITU-T Rec. H.263 offers several features that make it superior in performance to ITU-T Rec. H.261. In its baseline mode, ITU-T Rec. H.263 offers half-pel motion compensation, unlike the full-pel motion compensation of H.261 which is augmented by a loop filter. The baseline mode also includes several other aspects which can improve its efficiency (including 3-D variable-length coding, median motion vector prediction, deletable GOB headers, and more efficient coding pattern signalling). ITU-T Rec. H.263 also offers a wide variety of optional modes which may be supported at the discretion of the codec implementation. Each of these modes can be added to the baseline algorithm to improve the coding performance or to broaden its application range in various ways. These options include:

- features for enhancing perceptual quality and improving coding efficiency;
- features for altering the picture resolution in an adaptive manner;
- features for enhancing resilience to channel errors and packet losses (when interfacing to some network environments);
- features for providing bitstream scalability for error-prone and heterogeneous-network environments;
- the ability to add supplemental enhancement information for use with the video data.

In addition to the QCIF and CIF formats found in ITU-T Rec. H.261, ITU-T Rec. H.263 supports the following standard source formats:

- SQCIF  $128 \times 96 \times 29.97$  fps with 12:11 pixel aspect ratio and 16:11 picture aspect ratio;
- 4CIF  $704 \times 576 \times 29.97$  fps with 12:11 pixel aspect ratio and 16:11 picture aspect ratio;
- 16CIF  $1408 \times 1152 \times 29.97$  fps with 12:11 pixel aspect ratio and 16:11 picture aspect ratio.

A wide variety of custom picture sizes, custom picture clock frequencies, and custom pixel aspect ratios (and therefore custom picture aspect ratios) are also supported in ITU-T Rec. H.263. Still image transmission using the format specified in Annex D/H.261 is not an allowable mode of ITU-T Rec. H.263. Instead, high-resolution formats such as the 4CIF format of Annex D/H.261 can be used directly as supported video formats, and picture snapshot tagging can be used to indicate the presence of still image snapshots within the ITU-T Rec. H.263 video stream. Progressive refinement of such pictures can also be supported by progressive refinement tagging.

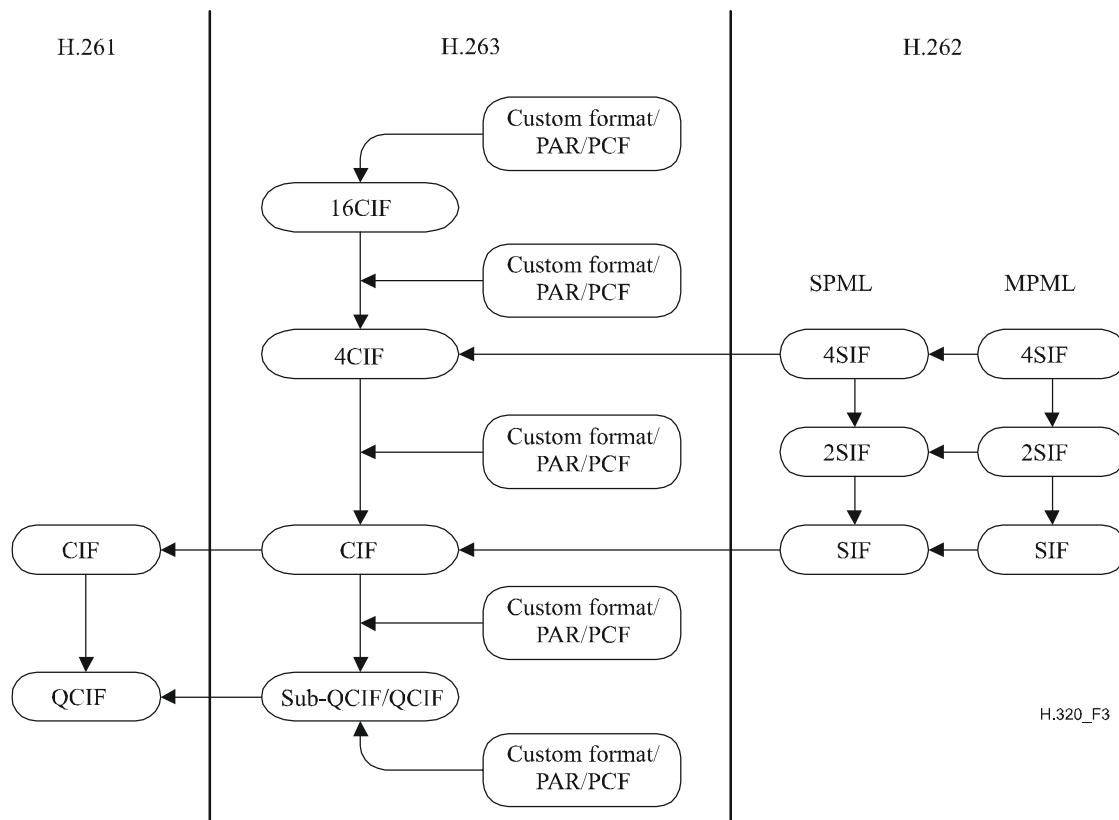
H.263 codecs for H.320 shall support the same capabilities for the sub-QCIF (SQCIF) and QCIF image sizes. This establishes the minimum quality at the QCIF image size, since this is the smallest image size supported by H.261, and the purpose of the optional video codecs is to improve upon H.261's performance.

H.263 codecs in H.320 systems shall support BCH error correction coding. (For details on BCH error correction, refer to 5.4/H.261 and Annex H/H.263.) When H.263 is used in H.320, BCH error correction is mandatory for both encoders and decoders.

#### A.4 Enhanced video algorithm hierarchy

To enforce a guaranteed level of interoperability among enhanced H.320 video codecs, a hierarchy has been created for enhanced video codec implementation. As H.261 is present in all current video-capable H.320 systems, H.261 is mandatory in any enhanced H.320 system with video capability. As H.263 should provide improved performance over H.261 at all times, baseline H.263 capability shall be required in systems that use enhanced video modes. Once H.261 and baseline H.263 have been implemented, H.262 may also be utilized. If, in the future, a new video algorithm is developed, then this hierarchy will need to be modified to introduce another layer or branch dependent on the algorithm's relative performance, range of application, or other factors.

The hierarchy is resolution-based, so a system must only support the equivalent resolution (and any lower resolutions) of the lower algorithms in the hierarchy when supporting a higher level algorithm. (In this hierarchical system, H.261 is considered the lowest level algorithm, H.262 MPML the highest. H.261\_CIF/H.263\_CIF and H.263\_4CIF/H.262\_4SIF are examples of equivalent resolutions.) The hierarchy also mandates that a lower level algorithm must have the same or better MPI capability as a higher level algorithm at an equivalent resolution. (A better MPI capability means that the decoder can use a lower MPI value, thus supporting a higher frame rate.) For example, a system designer wishing to implement H.262 MPML SIF capability must also implement H.262 SPML SIF, H.263 CIF, H.263 QCIF, H.263 SQCIF, H.261 CIF and H.261 QCIF. In the implementation, the MPI capability for H.262 SPML SIF must be better than or equal to the H.262 MPML SIF MPI, the H.263 CIF MPI capability must be better than or equal to the H.262 SPML SIF MPI capability, and the H.261 CIF MPI capability must be better than or equal to the H.263 CIF MPI capability. In addition, the H.261 QCIF MPI capability must be better than or equal to the H.263 QCIF/SQCIF MPI capability. The full video hierarchy is shown in Figure A.1.



NOTE – Arrows point to source formats which must be supported before the current source format.

**Figure A.1/H.320 – Mandatory video hierarchy for enhanced H.320 video implementation**

H.263 can use custom picture sizes, custom pixel aspect ratios and custom picture clock frequencies. The hierarchy rules when any of these are used are given below:

Any coder with a custom picture size shall support the "equivalent" standard picture size, where the equivalent size is the largest standard picture size smaller than or equal to the upper bound of the resolution range specified in the custom picture size in both height and width. If a custom picture size is smaller than QCIF, the equivalent resolution will be QCIF. For example, if the custom picture size range is  $[176-528] \times [144-432]$ , the equivalent standard size is CIF ( $352 \times 288$ ). If the custom picture size is  $120 \times 90$ , the equivalent standard size is QCIF ( $176 \times 144$ ).

If a H.320 terminal supports a custom pixel aspect ratio for any standard picture size, it shall also support the standard pixel aspect ratio (12:11) for that picture size. If a H.320 terminal supports a custom pixel aspect ratio for a custom picture size, it shall also support the standard pixel aspect ratio (12:11) for the equivalent standard picture size.

If a H.320 terminal supports a minimum picture interval (MPI) of greater than or equal to  $1001/30\ 000$  seconds with a custom picture clock frequency for any standard picture size, it shall also support a better (i.e., smaller) or equal MPI for all smaller standard picture sizes at the standard picture clock frequency ( $30\ 000/1001$  Hz, approximately 29.97 Hz) where MPI is measured in seconds (and not picture clock ticks). If a H.320 terminal supports a MPI of less than  $1001/30\ 000$  seconds (necessarily with a custom picture clock frequency) for any standard picture size, it shall also support a MPI of  $1001/30\ 000$  seconds for all smaller standard picture sizes at the standard picture clock frequency (i.e., the terminal shall support a MPI of 1 picture clock tick at the standard picture clock frequency). If a H.320 terminal supports a MPI of greater than or equal to  $1001/30\ 000$  seconds with a custom picture clock frequency for a custom picture format, it shall also support a better (i.e., smaller) or equal MPI for the equivalent standard picture size and all smaller standard picture sizes at the standard picture clock frequency. If a H.320 terminal supports a

MPI of less than 1001/30 000 seconds (necessarily with a custom picture clock frequency) for a custom picture format, it shall also support a MPI of 1001/30 000 seconds for the equivalent standard picture size and all smaller standard picture sizes at the standard picture clock frequency (i.e., the terminal shall support a MPI of 1 picture clock tick at the standard picture clock frequency).

Although the hierarchy across algorithms is strict in implementation, it is not automatically assumed during capabilities exchange. Capabilities must be stated independently for H.261, H.262 and H.263. For example, a system which signals that it is capable of H.262 SPML SIF with an MPI of 2 is strictly required to support the same or lower MPI value for H.263 CIF and H.261 CIF. However, separate codewords must be transmitted to declare the H.263 and H.261 capabilities.

ITU-T Rec. H.242 details the syntax and procedures for the use of the three algorithms, including mandatory stipulations regarding their individual format hierarchies and MPI values.

Asymmetric transmission of video algorithms is required in enhanced H.320 terminals. For example, a terminal may transmit H.262 SPML 4SIF while receiving H.261 CIF. (See 6.1/H.242 and clause 4/H.243 for more details on asymmetric video operation for point-to-point and multipoint use).

#### A.5 Use of ITU-T Rec. H.264 in H.320 systems

ITU-T Rec. H.264 represents an evolution of the existing video coding standards (H.261, H.262 and H.263) and it was developed in response to the growing need for higher compression of moving pictures for various applications such as videoconferencing, digital storage media, television broadcasting, Internet streaming, and communication.

The signalling and procedures for the use of H.264 in H.320 systems are defined in ITU-T Rec. H.241.



## **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, Internet protocol aspects and Next Generation Networks
- Series Z Languages and general software aspects for telecommunication systems