International Telegraph and Telephone
Consultative Committee
(CCITT)
Period 1985-1988

Document # 341
February 1988
Original: English

QUESTION: XV/4 Specialists Group/SGXV

STUDY GROUP XV - CONTRIBUTION

SOURCE: United States

TITLE: M x 56/64 kbps Video Telephony: Frame Format, Picture Format, Coding and Network Aspects leading to compatible standard.

1. Scope

This proposal provides a detailed description of the basic frame format, picture format, and coding algorithms for m x 64 kbps video telephone signals. These specifications will be implemented in equipment designed to terminate the digital transmission systems used to transport such signals (including their associated audio and data signals). For applications within Public Switched Telephone Networks, m will usually have a value of 1 or 2. (m) may also have a value of 3 through 6 (inclusive) for wider bandwidth applications.

2. VIDEO SOURCE FORMAT

The format to be coded is non-interlaced with an aspect ratio of 4:3. Pictures are coded in component form, these components being luminance (Y) and two color difference signals (CR, CB). The three components and the codes representing their sampled values are as defined in CCIR Rec. 601.

Black = 16
White = 235
Zero color difference = 128
Peak color difference = 16 and 240

Codes outside the above ranges will be accepted but may be modified by the coder to avoid emulation of reserved codewords.

Pictures are encoded at a maximum frequency negotiated at the start of the session.

Considering the following agreed to at the Tokyo meeting:

1. Realization of full CIF at mx64 kbit/s is the long term performance target.

2. Nevertheless full CIF operation is not envisaged easy to achieve at the nearest future with reasonable cost.

Two picture formats are described in the paper: a required format and an optional format. The format to be used during a session is negotiated at the start of the session. The optional format will be used if it is supported by both parties. Otherwise, the required format will be used.

Luminance and color differences are sampled in an orthogonal sampling arrangement. The color differences are subsampled 4:1 horizontally and 2:1 vertically (See Note 1). The picture formats are summarized in Table 1.

Table 1. Luminance and color difference sampling parameters

		Required format	Optional Format
Luminance	Pels per line	180	256
	Lines per picture	144	240*
Chrominance	Pels per line	45	64
	Lines per picture (1)	72	120*

Note 1: The color difference signals could be subsampled 4:1 in the vertical direction if further experiments demonstrate satisfactory performance.

*Note 2: It was noted that the optional format for lines/picture may be modified to be 192 for luminance and 96 for chrominance.

Note 3: The possibility of using higher resolution formats than those shown in Table 1 for stillframe transmission is for further study.

3. VIDEO SOURCE CODER (For Further Study)

4. AUDIO SOURCE CODER

The essential format for coding the associated audio signals shall conform to CCITT recommendation G.722. Other formats that can be negotiated within a session are under study.

5. NETWORK ASPECTS

5.1 Hypothetical Reference Connections

5.1.1 1 x 56 kbps, V.35 Interface:

Figure 1 shows a reference connection of two 56 kbps video telephones connected by a 56 kbps data service using V.35 interfaces. A 56 kHz clock is available at the V.35 interface, but a 8 kHz clock is not assumed. Possible network services include 56 kbps DDS or 56 kbps PSDS using conventional data terminals having the V.35 interface.

5.1.2 2 x 56 kbps, V.35 Interfaces:

Figure 2 shows a reference connection of two video telephone connected by two separate 56 kbps data services. Neither octet timing nor octet alignment is assumed. That is, there is no 8 kHz clock, and the absolute delay of the two 56 kbps connections may be different.

5.1.3 1 x 64 kbps with Octet Timing:

Figure 3 shows a reference connection of two video telephones connected by a 64 kbps network service. The network provides an 8 kHz clock to each terminal, and the octet transmitted during each 125 microsecond interval will be received intact during a 125 microsecond interval at the other end of the network connection. An example of a network service would be an ISDN 64 kbps unrestricted bearer capability call.

5.1.4 2 x 64 kbps, with Octet Timing, without Octet Alignment:

Figure 4 shows a reference connection consisting of two separate 64 kbps connections. In this case, there is octet timing as described previously. However, since the two connections were separately configured, there may be a difference in the absolute delay of the two connections. An example of a network service would be two separate 64 kbps unrestricted ISDN calls.

5.1.5 2, 3, 4, 5, or 6 xx 64 kbps, with Octet Timing and Alignment:

Figure 5 shows a reference connection consisting of two video telephones connected by facilities operating at a multiple of 64 kbps. In this case, there is zero differential delay between the several 64 kbps channels. [An example of a network service would be a DS1-rate private line directly connected from CP-CP.]

5.2 COMPATIBILITY

5.2.1 Between Videotelephones:

When videotelephones are connected, they will operate according to one of the five hypothetical reference connections. Since a different network service, or a different number of 56/64 kbps channels must be specified at the origination of the connection, the terminals will know a priori the method chosen for interworking.

At the time of connection, the terminals will exchange capability indication, including whether they are able to operate over two or more channels. It must be a user option to determine whether one or two channel operation is used in a given call, so that the user may control the trade-off between video and audio quality versus transmission cost. Furthermore, the terminal should be capable of single channel operation on either port, and if one channel of a pair in use becomes disconnected, communication should proceed without interruption on the remaining channel.

Manufacturers may choose to implement a subset of the capabilities.

5.2.2 Between Vidotelephones and PCM Terminals:

Since PCM terminals and/or interfaces to the PSTN requires 64 kbps operation with octet timing, interworking with PCM terminals is only possible in configuration 3. However, in the case of configuration 1 on a PSDS service, the DSU may have a separate means of intercommunication with PCM terminals.

5.2.3 Between Videotelephones and 7 kHz Audio or Audio/Data Terminals:

Intercommunication between videotelephones and 7 kHz audio or audio/data terminals is possible in configuration 3, since the same framing and signal formats are used.

It is noted that 7 kHz audio or audio/data terminals may also, in the future, be specified for 56 kbps operation (configuration 1). However, since an 8 x 80 bit frame is proposed for videotelephone operation, the videotelephones would not be compatible with 7 kHz audio or audio/data terminals at the 56 kbps rate. The latter must use a 7 x 80 bit frame in order to send 2-bit high band and 5-bit low band audio codes at the 8000 Hz rate.

5.3 CALLING PROCEDURES

Terminals will be able to distinguish which hypothetical reference connection is in use by virtue of which loop interface and which call type is being used.

5.3.1 64 kbps ISDN Calling:

Calls between 64 kbps terminals will observe the procedures of Rec. G. 72Y.

5.3.2 56 kbps Calling:

Calls within PSDS will be completed according to the procedures applying to setting up a 56 kbps digital data connection. Once the connection is set up, the terminals will come into frame, and the procedures of Rec. G.72Y will be used.

5.3.3 Intercommunication Between ISDN and PSDS Terminals:

If an ISDN videotelephone connects with a PSDS terminal, the connection will be a 56 kbps data call, rate adapted to the 64 kbps bearer channel, so far as the ISDN terminal knows. The ISDN videotelephone may optionally be equipped to intercommunicate with such an arrangement, operating as though it were a V.35 56 kbps videotelephone connected to an ISDN V.35 terminal adaptor.

6. TRANSMISSION CODER

6.1 Framing Signal

The transmission shall be arranged in 80 octet frames as specified in CCITT draft Recommendation Av.221. Sixteen frames are grouped in a multiframe as specified in AV.221. The least significant bit of each octet shall be the service channel, the first eight bits of the service channel in each frame shall carry the frame alignment signal, and the second eight bits of the service channel shall carry the bitrate allocation signal as specified in AV.221.

6.2 TRANSMISSION FORMATS

6.2.1 56 kbps Operation:

In 56 kbps operation, the octets of each 8 x 80 bit frame will be transmitted in order, most significant bit first, as the 56 kbps rate. Each frame will be transmitted in 2/175 seconds. Octet timing will be recovered from the frame alignment signal as specified in AV.221.

It is noted that this approach uniformly scales the bitrate available for each signal transmitted in the frame, e.g., if 2 bits of voice are transmitted in each octet, then the voice bitrate is reduced from 16000 b/s to 14000 b/s. If transmission at 16000 b/s is required by the voice coding algorithm, then voice can be transmitted in 3 bits per octet, in 16 out of 21 submiltiframes. Alternatively, a variable rate voice coding algorithm may be specified, depending on the progress in voice coding standards.

6.2.2 2 x 56 kbit/s Operation:

In 2 x 56 operation, each of the 56 kbit/s channels will be framed and transmitted separately. Octet timing will be recovered independently from the frame alignment signal of each channel, and the differential delay between the channels will be compensated for on the basis of the multiframe numbering method specified in AV.221.

The voice signal will be carried in one of the two channels. However, video, graphics and auxiliary data communications may be carried by either or both of the channels.

6.2.3 64 kbps Operation:

In 64 kbps operation the octets of each frame will be transmitted in order, most significant bit first, at the 64 kbit/s rate. Each frame will be transmitted in 10 msec. Octet alignment will be derived from the network, and the terminal will use network octet timing to send correctly aligned octets. However, the terminal may choose to frame on the received signal independent of network octet timing.

6.2.4 2 x 64 kbps Operation:

Octet timing will be derived from the network, and the terminal will transmit correctly aligned octets.

6.2.4.1 Low Rate Encoded Speech:

When low rate encoded speech is used, then speech will be transmitted in one of the two channels. However, either or both of the channels may contain video, graphics, and auxiliary data signals.

Each channel will be framed separately, and differential delay will be determined and compensated for by the multiframe numbering technique of AV.221.

6.2.4.2 7 kHz Audio:

When 7 kHz audio is used, then one of the two 64 kbps channels will contain video, graphics, and auxiliary data signals.

The other channel will contain only audio and auxiliary data signals. This channel will be framed and transmitted according to AV.221 and G.72y.

Differential delay between the two channels need not be compensated for, considering that it will be possible to limit differential delay to 50 msec by ISDN signaling requests, or by private network design.

6.2.5 2, 3, 4, 5 or 6 x 64 kbps

6.2.5.1 With Low Rate Speech Encoding:

In this case the signal is structured into 80 octet frames and the octets are transmitted in order, most significant bit first, at the rate of 128, 192, 256, 320 or 384 kbps. The time to transmit one frame is 5, 10/3, 2.5, or 5/3 ms.

As in the case of 56 kbps transmission, it is noted that this approach causes the transmission rate of each signal transmitted in the frame to scale proportionately. Thus, if 2 bits per frame are used for speech, the speech rate for 6 x 64 kbps transmission becomes 96000 bit/s. However, in this case the speech can be transmitted in every sixth two-frame submultiframe in order to reduce the average speech transmission rate to 16000 bit/s.

6.2.5.2 With 7 kHz Audio Encoding:

In this case 1, 2, 3, 4, or 5 channels are used for the transmission of video, graphic and auxiliary data information, and they are combined to form a framed channel of 64, 128, 192, 256 or 320 kbps.

The other channel is framed separately, and carries 7 kHz audio and auxiliary data channels as described for the 2 x 64 kbps case above.

The compatibility of 6 x 64 kbps with 384 kbps video telephony is for further study.

6.2.5.3 Dynamic Allocation within DS1 Rate Channels:

Intelligent terminals may have a means for dynamically increasing or decreasing the bitrate during a connection. The means for controlling such allocations are beyond the scope of this standard, since the means must operate on channels not carrying videotelephony information. However, it is noted that the 2, 3, 4, 5, and 6 x 64 kbit/s signals are compatible with this mode of operation, and that the low rate encoded speech format and the video/graphics part of the 7 kHz audio format can be sent at any speed, not necessarily multiples of 64 kbps. When rates which are not multiples of 64 are used, then network octet timing is not meaningful, and the framing must be extracted from the received signal independently.

6.3 BITRATE ALLOCATION

The bitrate allocation signal is transmitted according to AV.221, and defines the content of each two-frame submultiframe.

6.3.1 Signals to be Transmitted: Several different signals must be communicated as part of audio visual services. When there is no information to be transmitted we have:

N - Neutral: the bit position is assigned to carry no information.

Several signals are carried at relatively high bit rates and are mutually exclusive. Only one of the following can be carried in a 64 kbps channel at a given time:

P - PCM voice encoded according to A or mu-law G.711.

H,L - High and low band parts of 7 kHz audio encoded per G.722.

D - Data at 56 or 64 kbit/s.

V - Video

G - Graphic or other high speed audio/visual application data.

Signals which are compatible with, and can be transmitted at the same time as the "exclusive" signals are:

c - Command and indication information. The number of c bits per frame is for further study, and the protocol for command and indication information is to be defined.

a - Application channel data carried between AV or VP terminals according to standard protocols and recommendations to be defined.

d - Low speed data carried between Data Terminal Equipment (DTE) attached to the AV or VP terminals, without necessarily having any standard structure or protocol.

e - Low rate encoded speech at 8, 16, or 32 kbps. The speech encoding rate and algorithm are predetermined at connection origination.

6.3.2 Statistical Multiplexing:

When a channel is desired at other than the nominal rate available by continuously transmitting the signal in every submultiframe, a slower rate may be obtained by transmitting the signal in some proportion of the submultiframes. For example, if an application channel at 2400 b/s is desired, and the C&I channel is not active, then 288 bits of application channel information can be sent in three out of eight submultiframes.

No provision is made for the subdivision of application or data channels by means of BAS values. The subdivision may be done according to available data communications standards.

6.3.3 BITRATE ALLOCATION SIGNAL (BAS) ASSIGNMENTS

Table 2 defines 20 BAS values for Video/Speech and Table 3 defines 22 BAS for Audio/Data. The method for coding BAS values for Video /Speech is shown in Table 4. The encoding method of Table 4 leads to a straightforward design of the multiplexor, where simple logic can be used to determine which source is to be transmitted in each bit position, except for bit position 8. Bit position 8 already contains at least the FAS and BAS, so special logic which takes into account the position within the frame is already needed.

The layout for audio/data BAS in Table 3 is nearly as regular with respect to G.722 and 56 kbit/s data modes. However, the four unframed modes, and the neutral channel modes are placed in the group of 0x1xx codes, so that they may be detected easily, and then processed appropriately.

6.4 ERROR HANDLING

Under Study

7. COMMUNICATIONS PROCEDURE

A videophone session is divided into the following phases.

Call set-up phase (Phase A). This phase sets the transmission path between terminals.

Mode setting phase (Phase B). The capabilities of each terminal are exchanged, and parameters for the session are negotiated.

Information transfer phase (Phase C). The actual communication is performed.

Mode closing phase (Phase D). Opposite to phase B; returns the terminal to the state immediately after call set-up.

Call release phase (Phase E). The transmission path is disconnected.

7.1 MODE SETTING PHASE

The parameter negotiation procedure is divided into two parts. The first part (Phase B1) is defined in AV.221 and identifies the terminal as a videophone terminal by sending a specific BAS code. There are different BAS codes assigned for one- and two-line terminals.

During the second part (Phase B2), the terminals exchange the following information:

Manufacturer identification
Terminal identification
Terminal capabilities

Terminal capabilities include:

Picture format
Maximum frame rate
Video algorithm
Audio algorithm
Data ports (number of channels, maximum rates)
Graphics/cursor capabilities
Telewriter capability
Display features (split-screen, windowing)
Duplex/broadcast terminal
Encryption capability

When the capabilities have been exchanged, maximum common mode is selected for the session. The selected parameters are then communicated and acknowledged.

The detailed procedure for Phase B2 remains for further study.

7.2 OPERATION OVER 2 X 56/64 KBPS

When two channels are used for the videophone session, the call setup and mode setting phases are performed independently on each channel. The terminal identification provides a means of verifying that both calls originate from the same remote terminal.

The procedure for entering the information transfer phase remains for further study.

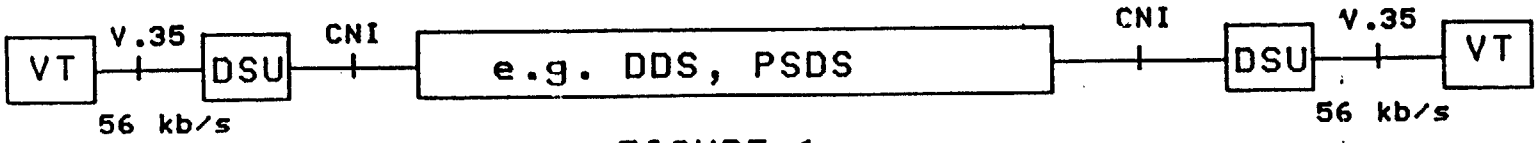

FIGURE 1

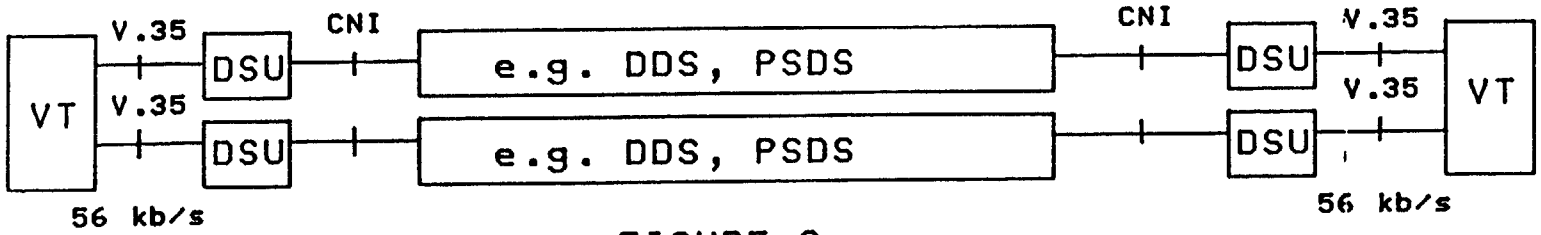

FIGURE 2

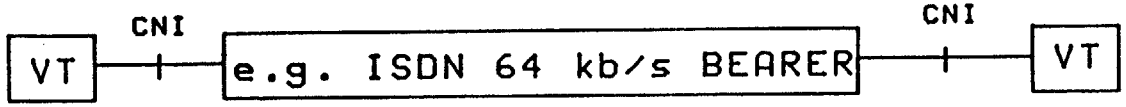

FIGURE 3

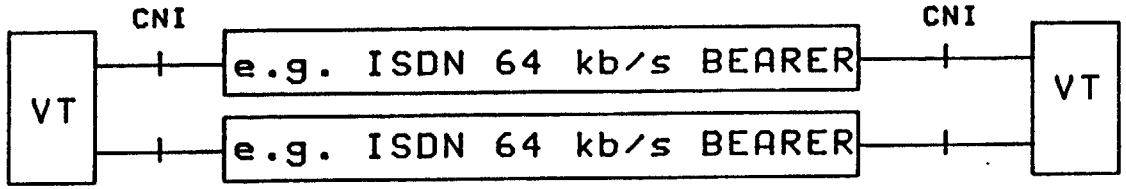

FIGURE 4

FIGURE 5

Attribute Value	Octet Layout								Bits per Submultiframe				
Bits 12-16	1	2	3	4	5	6	7	8	Video	Speech	Data	App.	C&I
00000	V	V	V	V	V	V	V	V	1248	0	0	0	0
00001	V	V	V	V	V	V	V	c/V	1248-C	0	0	0	C
00010	-	-	-	-	-	-	-	-	-	-	-	-	-
00011	-	-	-	-	-	-	-	-	-	-	-	-	-
00100	-	-	-	-	-	-	-	-	-	-	-	-	-
00101	-	-	-	-	-	-	-	-	-	-	-	-	-
00110	-	-	-	-	-	-	-	-	-	-	-	-	-
00111	-	-	-	-	-	-	-	-	-	-	-	-	-
01000	V	V	V	V	V	V	V	d	1120	0	128	0	0
01001	V	V	V	V	V	V	V	c/d	1120	0	128-C	0	C
01010	V	V	V	V	V	V	V	a	1120	0	0	128	0
01011	V	V	V	V	V	V	V	c/a	1120	0	0	128-C	C
01100	V	V	V	V	V	V	d	d	960	0	288	0	0
01101	V	V	V	V	V	V	d	c/d	960	0	288-C	0	C
01110	V	V	V	V	V	V	a	a	960	0	0	288	0
01111	V	V	V	V	V	V	a	c/a	960	0	0	288-C	C
10000	s	s/V	s/V	s/V	V	V	V	V	928	320	0	0	0
10001	s	s/V	s/V	s/V	V	V	V	c/V	928-C	320	0	0	C
10010	-	-	-	-	-	-	-	-	-	-	-	-	-
10011	-	-	-	-	-	-	-	-	-	-	-	-	-
10100	-	-	-	-	-	-	-	-	-	-	-	-	-
10101	-	-	-	-	-	-	-	-	-	-	-	-	-
10110	-	-	-	-	-	-	-	-	-	-	-	-	-
10111	-	-	-	-	-	-	-	-	-	-	-	-	-
11000	s	s/V	s/V	s/V	V	V	V	d	800	320	128	0	0
11001	s	s/V	s/V	s/V	V	V	V	c/d	800	320	128-C	0	C
11010	s	s/V	s/V	s/V	V	V	V	a	800	320	0	128	0
11011	s	s/V	s/V	s/V	V	V	V	c/a	800	320	0	128-C	C
11100	s	s/V	s/V	s/V	V	V	d	d	640	320	288	0	0
11101	s	s/V	s/V	s/V	V	V	d	c/d	640	320	288-C	0	C
11110	s	s/V	s/V	s/V	V	V	a	a	640	320	0	288	0
11111	s	s/V	s/V	s/V	V	V	a	c/a	640	320	0	288-C	C

Note: Video and speech bit rates are shown for 16 kb/s speech.

Table 1. Octet Formats and Attribute Value Assignments for Video/Speech Transmitted in 8 x 80 bit Frames.

Attribute Value	Octet Layout								Bits per Submultiframe				
Bits 12-16	1	2	3	4	5	6	7	8	D Data	Speech	d Data	App.	C&I
00000	D	D	D	D	D	D	D	N	1120	0	0	0	0
00001	D	D	D	D	D	D	D	c/N	1120	0	0	0	C
00010	-	-	-	-	-	-	-	-	-	-	-	-	-
00011	-	-	-	-	-	-	-	-	-	-	-	-	-
00100	P	P	P	P	P	P	P	P	0	1280A	0	0	0
00101	P	P	P	P	P	P	P	P	0	1280mu	0	0	0
00110	H	H	L	L	L	L	L	L	0	1280	0	0	0
00111	D	D	D	D	D	D	D	D	1280	0	0	0	0
01000	D	D	D	D	D	D	D	d	1120	0	128	0	0
01001	D	D	D	D	D	D	D	c/d	1120	0	128-C	0	C
01010	D	D	D	D	D	D	D	a	1120	0	0	128	0
01011	D	D	D	D	D	D	D	c/a	1120	0	0	128-C	C
01100	-	-	-	-	-	-	-	-	-	-	-	-	-
01101	-	-	-	-	-	-	-	-	-	-	-	-	-
01110	N	N	N	N	N	N	N	N	0	0	0	0	0
01111	N	N	N	N	N	N	N	c/N	0	0	0	0	C
10000	H	H	L	L	L	L	L	L	0	1248	0	0	0
10001	H	H	L	L	L	L	L	c/L	0	1248-C	0	0	C
10010	-	-	-	-	-	-	-	-	-	-	-	-	-
10011	-	-	-	-	-	-	-	-	-	-	-	-	-
10100	-	-	-	-	-	-	-	-	-	-	-	-	-
10101	-	-	-	-	-	-	-	-	-	-	-	-	-
10110	-	-	-	-	-	-	-	-	-	-	-	-	-
10111	-	-	-	-	-	-	-	-	-	-	-	-	-
11000	H	H	L	L	L	L	L	d	0	1120	128	0	0
11001	H	H	L	L	L	L	L	c/d	0	1120	128-C	0	C
11010	H	H	L	L	L	L	L	a	0	1120	0	128	0
11011	H	H	L	L	L	L	L	c/a	0	1120	0	128-C	C
11100	H	H	L	L	L	L	d	d	0	960	288	0	0
11101	H	H	L	L	L	L	d	c/d	0	960	288-C	0	C
11110	H	H	L	L	L	L	a	a	0	960	0	288	0
11111	H	H	L	L	L	L	a	c/a	0	960	0	288-C	C

Note: 00100, 00101, 00110, and 00111 are unframed modes.

Table 2. Octet Formats and Attribute Value Assignments for Audio/Data (000) Transmitted in 8 x 80 bit Frames.

Attribute Value Bit N	Bit N = 0	Bit N = 1
N = 12	speech off	speech on
13	d and a off	d or a on
14	d or a in bit 8	d or a in bits 7 and 8
15	data channel on	application channel on
16	C&I off	C&I on