CCITT SGXV
Working Party XV/1
Experts Group for ATM Video Coding
(Rapporteur's Group on Part of Q.3/XV)

Document AVC-419
January 1993

SOURCE: JAPAN
TITLE: USE OF COMPATIBILITY/SPATIAL SCALABILITY IN MULTIPOINT AUDIOVISUAL COMMUNICATION SYSTEMS
Purpose: Discussion

---

1. Introduction

"Profile" is currently being discussed in order to make MPEG-2|H.26X a generic standard. What CCITT aims at are specifications for the interoperable audiovisual communication system and terminal. From this viewpoint, we examine here the "compatibility/spatial scalability" toward clarifying its role in the video coding standard.

Document AVC-278 (Japan) for the New Jersey/Rio de Janeiro meeting analyzed in what situations the compatibility between H.26X and H.261 is needed and what provisions are necessary in the H.26X coder/decoder. The following was concluded (see appended illustrations);

1) *The decoder in B-ISDN should be able to receive both of H.26X bitstream and H.261 stream, but one at a time, in all the situations including point-to-point, multipoint, multicast and database access.*

2) *H.26X encoder should be able to transmit either of H.26X bitstream or .261 stream at a time for;*

   - *point-to-point communications,*
   - *multipoint communications where a common operating mode is used, or*
   - *multipoint communications where MCU carries out transcoding.*

3) *It should also be able to transmit two bit streams simultaneously for;*

   - *multipoint communications where MCU does not provide transcoding,*
   - *multipoint communications where a mesh connection is employed,*
   - *multicast communications, or*
   - *database.*

This contribution further analyzes the multipoint communication system where both of H.32X terminals (containing H.26X as video coding) and H.320 terminals (containing H.261 as video coding) are accommodated in a MCU. Switched video scheme is assumed to simplify the discussion.

2. Methods to achieve multipoint communications

The three methods shown in Figure 1 will be practical. Operations of the terminal and MCU are as follows;

1) Embedded

H.26X sends out an embedded bitstream using the compatibility/spatial scalability syntax.

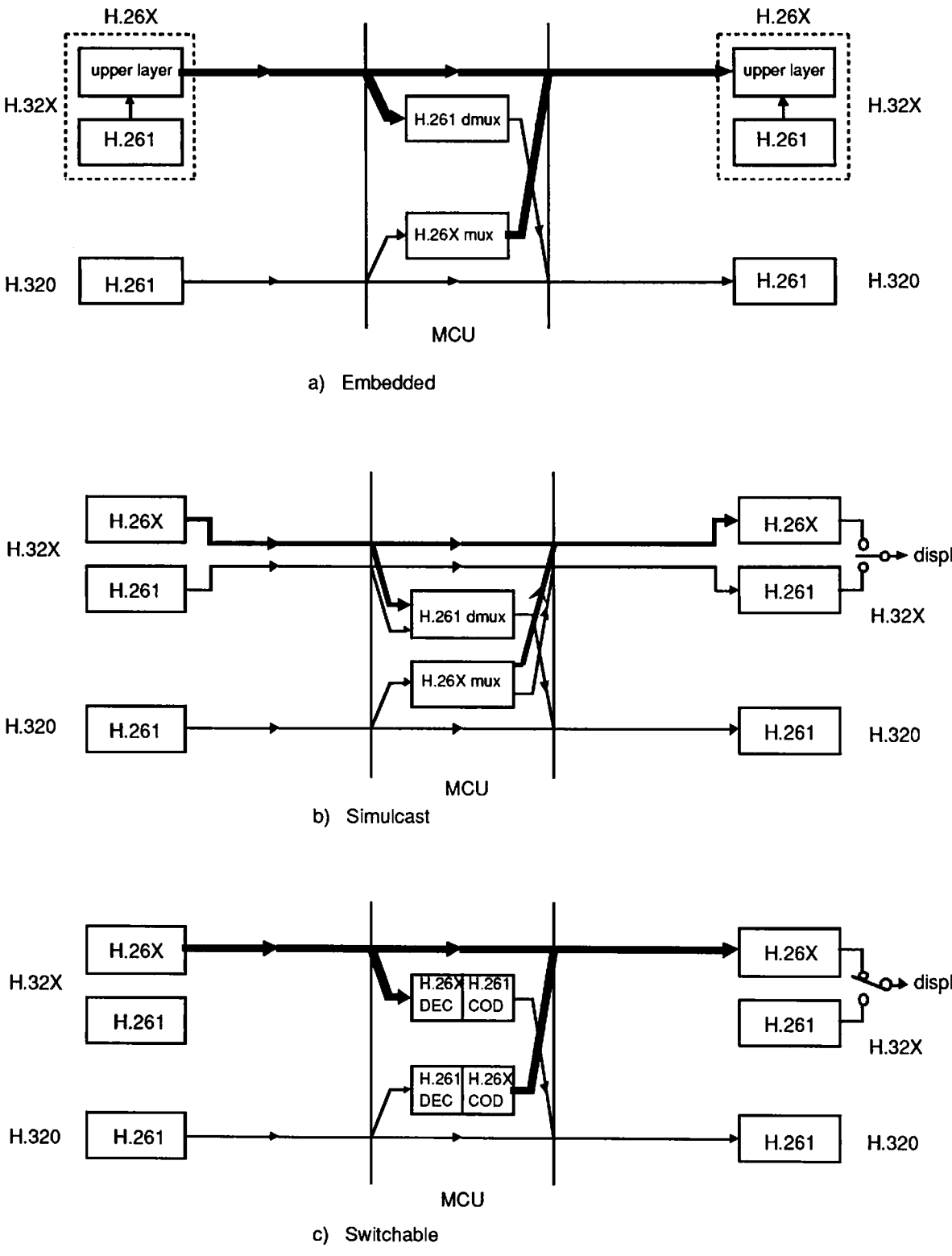


Figure 1 Multipoint visual telephone system involving both of H.32X and H.320 terminals

MCU operation

Video Source	MCU → H.32X	MCU → H.320
H.32X	H.26X embedded	H.261 extracted from the embedded H.26X
H.320	H.261	H.261

Features

+) Coding efficiency is high, particularly for the case bit rates of H.26X and H.261 are close like 4 and 1.5 Mbit/s.
+) Coding delay of H.26X does not degrade if high speed processor is used in a time division manner for base layer and upper layer coding. If an existing H.261 hardware is to be used for constituting embedded codec, delay for the upper layer coding is governed by H.261 delay characteristics.
+) For communications not involving H.261, the base layer coding can be switched off to minimize the coding efficiency loss.
-) Support of the compatibility/spatial scalability becomes mandatory for all the visual telephone terminals, thus increasing hardware burden to implement the embedded structure? Or is this additional hardware for the embedded structure is minimal because the H.32X terminal should be able to generate both of H.26X and H.261 bitstreams in any case?

2) Simulcast

H.32X sends out both of H.26X and H.261 bitstreams concurrently.

MCU operation

Video Source	MCU → H.32X	MCU → H.320
H.32X	H.26X as selected	H.261 as selected
H.320	H.261	H.261

Features

+) Multiplexing and demultiplexing of H.26X/H.261 at the terminal can be a simple processing.
+) All the H.32X terminals should have the H.320 emulation mode for interworking with H.320 terminals connected to N-ISDN, hence they should be able to switch H.26X and H.261. If separate hardware (chips) is used for this purpose, additional hardware to simultaneously operate them can be minimal? When single hardware is used in a time division manner, it becomes more complex compared to the simple switchable case?
-) Coding efficiency degrades compared to the embedded case, but the degradation can be minimal if the H.261 bit rate is low compared to that of H.26X.

3) Switchable

H.32X sends out only H.26X bitstream.

MCU operation

Video Source	MCU → H.32X	MCU → H.320
H.32X	H.26X without processing	H.261 transcoded from H.26X
H.320	H.26X transcoded from H.261	H.261 without processing

Features

+) H.32X terminal configuration is the simplest, but support of H.320 is mandatory.
-) Picture quality degrades and coding-decoding delay increases due to transcoding at MCU. When the H.26X bit rate is high, however, total picture quality and delay performance of transcoding to/from low bit rate H.261 is mostly determined by H.261.

3. Comparison

The above analysis is summarized in the following table indicating qualitative orders. Picture quality rows indicate also an example of H.32X using 4 Mbit/s and H.320 using 1.5 Mbit/s for video.

Item	Embedded	Simulcast	Switchable
Terminal implementation	3	2	1
H.32X → H.32X quality*	2 (H.26Xe, 4M)	3 (H.26Xs, 2.5M)	1 (H.26Xs, 4M)
H.32X → H.320 quality*	1 (H.261, 1.5M)	1 (H.261, 1.5M)	3 (H.26Xs, 4M) and (H.261, 1.5M) cascaded
H.320 → H.32X quality*	1 (H.261, 1.5M)	1 (H.261, 1.5M)	3 (H.261, 1.5M) and (H.26Xs, 4M) cascaded
H.32X → H.320 delay	1	1	3
H.320 → H.32X delay	1	1	3
Mux/dmux processing at MCU	3	2	1
Video processing at MCU	1	1	3

*) H.26Xe, 4M = H.26X embedded at 4 Mbit/s
H.26Xs, 4M = H.26X single layered at 4 Mbit/s
H.26Xs, 2.5M = H.26X single layered at 2.5 Mbit/s
H.261, 1.5M= H.261 at 1.5 Mbit/s

4. Items requiring further input

1) Bit rates

What are typical bit rates for the conversational service in B-ISDN and N-ISDN, respectively? Can they be 0.1 Mbit/s, 0.3 Mbit/s, 1 Mbit/s, 3 Mbit/s, or 10 Mbit/s? These values affect the choice of appropriate video coding for the multipoint system.

If the two bit rates are close, use of the highest common mode ("mode down" in AVC-278, H.261 in this case) can be the choice. If the two bit rates are largely different, transcoding may practically not degrade the total quality.

2) Implementation complexity

What hardware complexity is expected for the terminal having different configuration; embedded, simulcast and switchable?

A first order approximation can be the number of pels per picture. "Embedded" and "simulcast" require 20 -30% more hardware than "switchable".

3) Multimedia multiplexing

For "embedded" or "simulcast', its bitstream contains two video signals. Are they multiplexed at the multimedia multiplex layer or the video multiplex layer? It is noted that the current

H.221 does not support multiple video nor multiple audio and that there is no mechanism to explicitly designate the video bit rate.

It is unknown whether more than two video signals are required for conversational services. Currently still picture is considered as a visual aid in videoconferencing. Its extension to moving pictures may be one example?

5. Conclusion

How to use the generic video coding standard is specified by the terminal Recommendation. For conversational services, the video codec contained in the H.32X terminal is required as follows;

1) It should be switchable between H.26X and H.261.
2) If the necessity is agreed, it should further support "embedded" or "simulcast" functionality either
   - as a mandatory specification, or
   - as an optional specification.

In casethat the optional specification solution is chosen, the item should be included in the negotiation between MCU and terminal.

Overall evaluation should be given to terminal implementation complexity, MCU complexity and characteristics regarding picture quality and delay.

END

Appendix 1 (from AVC-278)

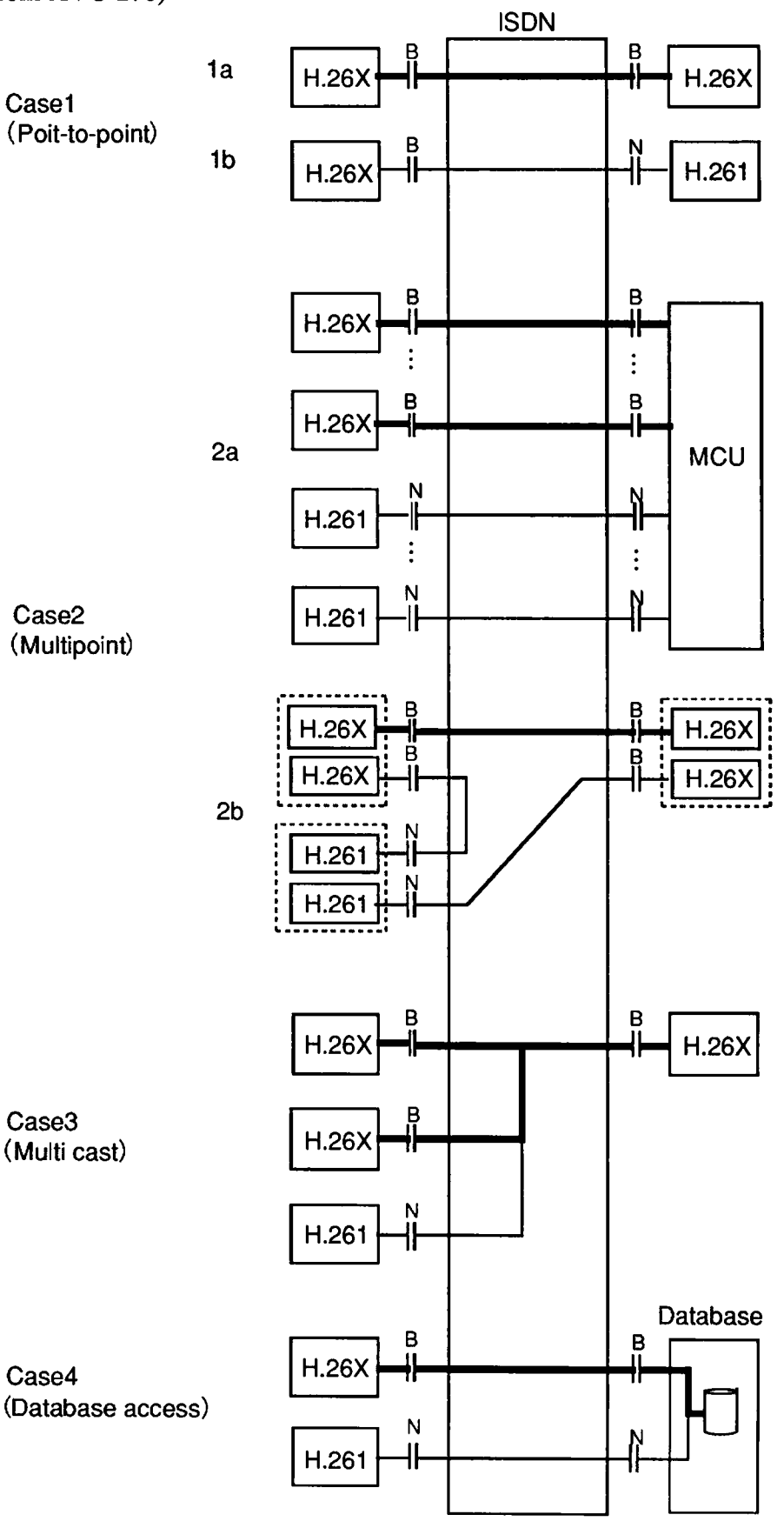
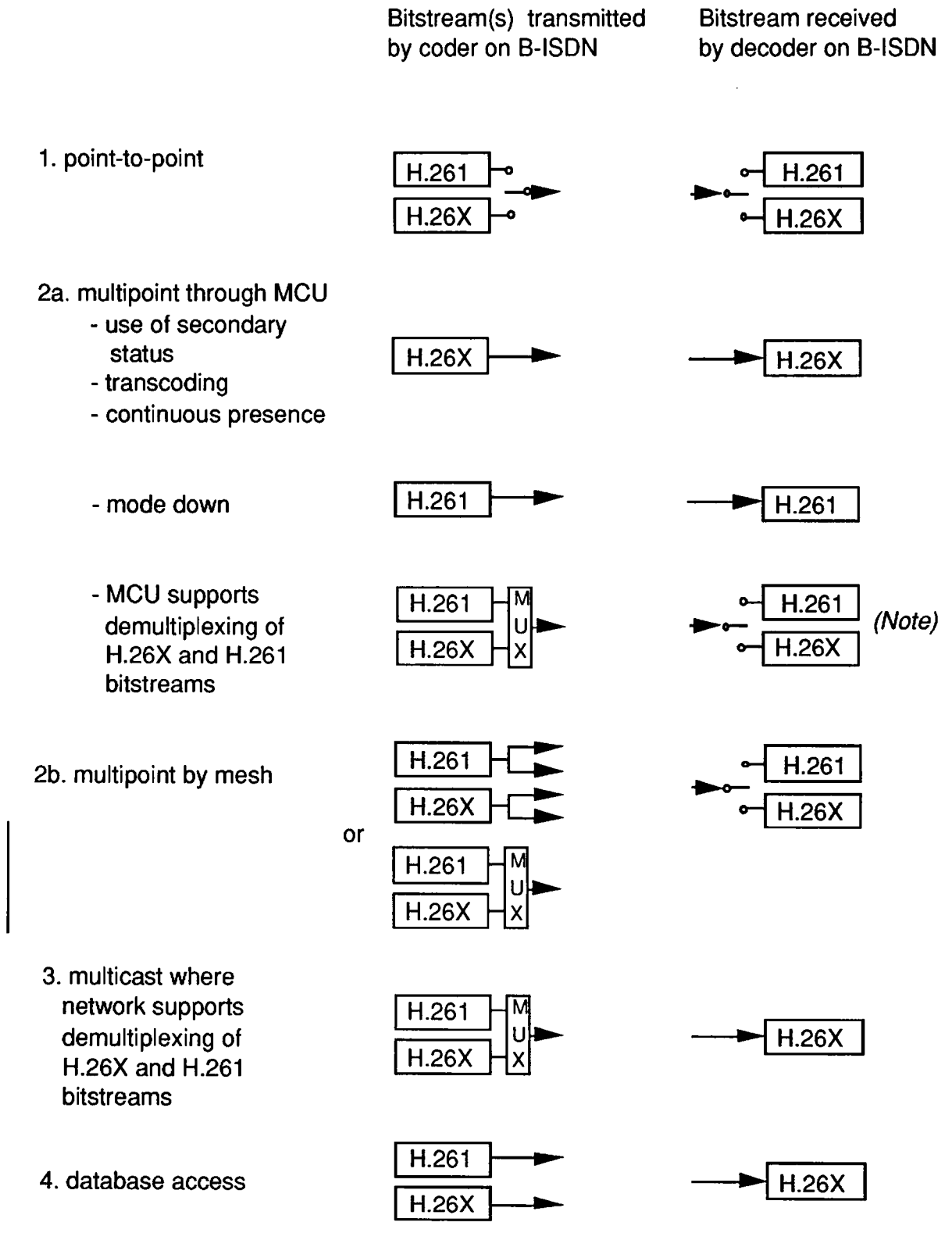


Figure1 Interworking between H.26X and H.261

Appendix 2 (from AVC-278 with addition as indicated by the bar at left)



Note : Fast switching may be required to respond to the change of senders.

Figure 2 Summary of H.261 compatibility requirements to the coder and decoder on B-ISDN