ITU Telecommunication Standardization Sector
Study Group 15
Experts Group for ATM Video Coding
(Rapporteur's Group on Part of Q.2/15)

Document AVC-560
September 2, 1993

Source: Bellcore
Title: Some Comments on Video Combining from 4QCIFs to CIF
Purpose: Information and discussion

I. Delay and HRD Tradeoff

We assume that

1. The bandwidth for the channel from encoder to MCU is 4R, while the averaged coded bitrate is R, based on symmetrical channel consideration. (Stuffing bits may be necessary.)
2. The MCU does not have to receive a complete QCIF frame before it transmits the corresponding CIF frame.

Upon receiving an intraframe (13 units for convenience), the MCU can do the following:

Case A:

```
Time In        a                b               c               d
     --------------------------------------------------------------------------------
IN1  |  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1
     --------------------------------------------------------------------------------
IN2  |  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1
     --------------------------------------------------------------------------------
IN3  |  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1|  1
     --------------------------------------------------------------------------------
IN4  |  1|13         |                                       |  1|  1|  1|  1|  1|  1
     --------------------------------------------------------------------------------
OUT  |  4|16             |3 |3 |3 |3 |3 |3 |3 |3 |3 |3 |3 |3 |  4|  4|  4|  4|  4|  4
     --------------------------------------------------------------------------------
Time Out                  a'          b'          c'            d'
```

In this case, the MCU does not add delay, but it generates a variable length output. The receiving terminal needs to buffer the variable length data (3 undecoded frames at time d'). This data would violate the HRD if the intraframe exceeded 16 units.

Case B:

```
Time In              a           b           c  d
    -------------------------------------------------------------
IN1 | 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1
    -------------------------------------------------------------
IN2 | 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1
    -------------------------------------------------------------
IN3 | 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1| 1
    -------------------------------------------------------------
IN4 | 1|13       |                            | 1| 1| 1| 1| 1| 1
    -------------------------------------------------------------
OUT | 4|3*|3*|3*|16         |3|3|3|3 |3|3|3|3 |3| 4| 4| 4| 4| 4| 4
    -------------------------------------------------------------
Time Out                     a'       b'       c' d'
                                                 3/4 frame time delay
```

In this case, the MCU adds 3/4 frame time delay, but the receiving terminal only has 2 undecoded frames at time d'.

Case A and Case B show a tradeoff between adding more MCU delay and violating the HRD sooner. An MCU will probably be designed for some maximum delay, and then pass the burden to the receiving terminal.

To meet the HRD requirement, one can reduce maximum QCIF frame bits (e.g. to 16 Kbits), and introduce a skip frame command. To skip the 3 extra frames in the Case A example, the MCU needs to send 3 skip frame commands to IN1, IN2, and IN3. Assuming everything works, the result may look like:

```
Time In     a           b     x     c     y     d
    -------------------------------------------------------------
IN1 | 1| 1| 1| 1| 0| 2| 0| 2| 1| 1| 1| 0| 2| 1| 1| 1| 1| 1| 1| 1
    -------------------------------------------------------------
IN2 | 1| 1| 1| 1| 0| 2| 1| 0| 2| 1| 1| 1| 0| 2| 1| 1| 1| 1| 1| 1
    -------------------------------------------------------------
IN3 | 1| 1| 1| 1| 0| 2| 1| 1| 0| 2| 1| 0| 2| 1| 1| 1| 1| 1| 1| 1
    -------------------------------------------------------------
IN4 | 1|13       |                            | 1| 1| 1| 1| 1| 1
    -------------------------------------------------------------
OUT | 4|16         |3|3 | 6 | 4| 4| 4|3| 5 | 4| 4| 4| 4| 4| 4| 4
    -------------------------------------------------------------
Time Out            a'        b' x'   c' y'     d'
```

All of these proposals imply changing the system requirements in H.320 and/or video encoding mechanism (for skipped frames), which need to be weighed against the MCU cost savings. However, any video decoder should be able to decode the mixer output.

It is true that even if each individual QCIF input to MCU meets its own HRD requirements, the combined CIF bitstream may violate the HRD conditions as a whole. However, whether such a violation is fatal to decoders is an issue. We feel that as long as the buffer content at the time when bits are taken out from the buffer, is smaller than

$$B + 256k - \int_{t_N}^{t_{N+1}} R(t)dt,$$

the violation is tolerable for the decoder. If the buffer size implemented in the decoder is greater than B+256k, violation tolerance increases. In case that the HRD does not satisfy the above relationship, the MCU can send a fast updating signal (specified in H.221) to each QCIF encoder for a replenishment. Since the combiner has all the coded data information, the MCU can duplicate a CIF HRD by itself. If its HRD is about to hit the B+256K bound, the MCU can clear its buffer and inform each QCIF client to fast update through H.221's bit-rate allocation signal (BAS). Upon receiving this fast update request, each QCIF client will "encode its next picture in INTRA mode with coding parameters such as to avoid buffer overflow" (H.261, section 4.3.2). By this way, HRD buffer overflow is prevented.

II. Frame synchronization

In the H.261's definition of temporal reference (sec. 4.2.1), it defines: "...TR is a 5 bit number which can have 32 possible values. It is formed by incrementing its value in the previously transmitted picture header by 1 plus the number of non-transmitted pictures (at 29.97 Hz) since that last transmitted one...." Also in sec. 3.1, "...Means shall be provided to restrict the maximum picture rate of encoders by having at least 0,1,2,or 3 non-transmitted pictures between transmitted ones. ..."

Accordingly, for a 15 frames/sec selection by H.221,

1. TR = 0,2,4,7,9,12,..... is a legitimate coding, but
2. TR = 0,2,3,7,9,12,,,, is not a legitimate coding because $3 - 2 = 1 < 2$.
3. TR = 30,0,4,8,12,... is legible.

In a multi-point videoconferencing, the MCU first has to poll the four coders for their frame rate capability. The minimum available frame rate will be adopted and mandated by H.221. To achieve frame synchronization between different terminals, the MCU can do the following two procedures in deciding the combining order:

(A) Convert each TR by first resetting the first available TR to "0" and then the subsequent TRs are circularly shifted by the same amount, e.g. the third sequence above is converted from TR = 30,0,4,8,12,... to TR = 0,2,6,10,14,....

(B) Do the following mapping (the mapping is unique following the above H.261 definition on TR and a frame rate of 15 frames/sec):

If the input QCIF's TR is	the combined CIF's TR is
0,1	1
2,3	3
4,5	5
...etc.	

Skipped QCIF frames' GOB headers need to be created. The maximum allowable CIF frame rate is thus the same as the specified maximum frame rate and all input frames will be combined appropriately within a fixed delay time.

Note that the MCU cannot simply group the available QCIF's together in their input order because input sequences' complexities are not the same and the inputs with less complexities or more frames may eventually overflow if some measures like the above TR mapping is not maintained.

III. Video Combining or Transcoding

When a new service capability is introduced, the question of whether existing terminals can be accommodated does not have a simple answer. For example, the current H.231/H.243 MCU requires symmetric video and data channels. Terminals that cannot support the communication mode selected by the MCU are put in secondary status, which generally means that they can participate by audio only. Therefore, the quality of service that can be provided to an existing terminal is debatable.

The 4QCIF to CIF video mixing contribution basically proposes a new optional video transmission mode, which is like any other optional audio, video, or data transmission mode. However, the objective is for the MCU to produce a CIF bit stream that can be decoded by an existing terminal. Therefore, in a conference involving old and new terminals, the old terminals can at least see the mixed video, and will not be made to work in audio only. How an MCU can do this is the subject of this discussion.

Of course, an MCU capable of transcoding audio and video may be fully transparent to all terminals (H.243 MCUs cannot be considered fully transparent). One reason for recommending the video switching MCU is because it is easier to implement than a video transcoding MCU. The purpose of the video mixing contribution is to examine if split screen video combining can be implemented in an acceptable way that is simpler than transcoding.

It is well-known that transcoding degrades picture quality and increases delay. To what extent the performance is degraded is an important study item, as suggested in AVC-553R. In addition, the cost and system complexity comparison between coded domain combining and transcoding which requires 4 QCIF decoders and 1 CIF encoder also need to be studied carefully.