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

Document AVC-581
November 2, 1993

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

Abstract

This contribution includes the following three main topics:

1. We describe two approaches which can handle asymmetrical video channels in a coded domain combiner (hereafter referred to as combiner) and a prototype implementing one of them.
2. We show that a combiner can be implemented without violating H.320. We also clarified some misunderstandings about our previous proposal AVC-501, as summarized in TD-43.
3. We believe that a combiner can co-exist with a pel domain transcoder (hereafter referred to as transcoder) since both can be implemented without violating H.320.

I. The first Approach to Implement An Asymmetrical Video Rate

In this approach, as proposed in AVC-500, full-rate is used for both up-link and down-link, and the asymetrical video rate is achieved by error correction frame bit-stuffing (as specified in H.261) at the transmitting side of each user. A possible advantage of this scheme is that since the full rate is used for the up-link, the end-to-end delay can be reduced. However, as shown in AVC-560 (September 2, 1993), the Hypothetical Reference Decoder (HRD) in H.261 may be violated in some cases using this approach. As explained in AVC-560, the violation of the HRD does not mean that decoder buffer will overflow, and it is not a fatal violation for many practical cases. If this violation is really not tolerable, the following approach can be used.

II. The Second Approach to Implement An Asymmetrical Video Rate

The second approach involves using a quarter rate up-link channel instead of a full rate up-link channel. The equivalent of a quarter rate up-link channel can be implemented in several ways. One way is to define the quarter rate channel through H.221/H.242. Note that in H.242 Sec.3.2, it is explicitly stated that "the H.242 procedures are equally applicable to systems in which asymmetric bidirectional communication is optimal (examples are surveillance and retrieval services)."

Similar to the figures presented in AVC-560, a figure of the combiner input/output for this approach is shown as follows. In the figure, the numbers indicate the units of compressed bits in a QCIF or CIF frame.

"3*" indicates 3 units of compressed bits plus 3 empty GOBs (with only GOB headers) and about 1 unit of fill bits. The assumptions in the figure (such as all the input QCIFs are synchronized in picture phase) are only used to simplify the illustration and are not needed for the actual operation.

```
     --------------------------------------------------------------------------------
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*| 3*| 3*| 3*| 3*| 3*| 3*| 3*| 3*| 3*|      16       |  4|  4|  4
     --------------------------------------------------------------------------------
```

Since only quarter rate is used in up-link video transmission, the delay in this approach is longer than the first approach. The advantages of this approach are that it is fully conforming to the H.320 standard and that HRD is not violated. In either approach, the delay is determined by the largest frame and will not accumulate.

III. A Simple Combiner Prototype and Some Clarifications to TD-43

Using desktop video conferencing (DVC) boards made by Integrated Information Technology (IIT), Ethernet cards made by 3COM, and a 486/DX2-66 personal computer (PC), we were able to implement the second approach in real-time by simple PC software.

A major task in this prototype is the implementation of the video combiner software to achieve real-time processing. We have successfully combined bit streams coded by IIT, Hitachi, VideoTelecom and our own H.261 software. Since we cannot find inexpensive and easily adaptable ISDN equipment for this prototype, we used the simpler and cheaper Ethernet. A very small network buffer and a modest video combiner buffer with a size comparable to the analysis of AVC-501 were selected. The real-time prototype has been demonstrated.

Prior to the prototype implementation we have done simplified analysis and simulations as shown in AVC-501 to investigate its feasibility. Unfortunately, several simplified assumptions used to ease simulation in AVC-501 caused a lot of misunderstandings. Most issues have been summarized in the TD-43 of the Geneva meeting in September 1993. Several issues are clarified as follows:

1. In TD-43, sec. 6.1, it stated the combiner can only operate properly when picture frequencies are locked to each other, coded rates are exactly the same and constant, and number of bits per QCIF frame are the same. Actually, this is only the assumption we used to derive the minimum achievable delay and buffer size (AVC-501, sec. 4.2.1). In our simulation, we have used real coded data and a

FIFO simulation (AVC-501 sec. 5). In the prototype, two real-time inputs are combined with two pre-coded sequences from different sources. We also stress-tested the system by blackening out one or two camera inputs frequently and by moving fast or stand still, in order to generate un-even intra or inter coded frame rates and skipped frames. The system works properly without any problems.

2. In TD-43, sec. 6.2 stated that picture inputs are not synchronized and their frame rates may vary, which might result in sophisticated management mechanism and large buffers in implementation. Synchronized picture input is only a simplification used in the simulation of AVC-501. The synchronization issue has been discussed in AVC-560. In the prototype, the system is able to robustly handle asynchronous picture inputs with widely varying frame rates. Sophisticated management mechanism and large buffers are not needed.

3. In TD-43, sec. 6.3, some figures were cited from AVC-501 to argue that a combiner gains only marginal advantage over a transcoder in terms of delay. We believe that the figured quoted were not based on a fair comparison. According to our simulation in AVC-501, in the best case, the transcoder delay is 352 ms, which is 86 ms longer than the combiner delay of 266 ms. In the worst case, the transcoding delay is 792 ms, which is 148 ms longer than the combiner delay of 644 ms. In average, the transcoding delay is 572 ms, which is 132 ms longer than the combiner delay of 440 ms. Note that both transcoder and combiner have to account for frame synchronization delay.

4. In TD-43, sec. 6.4, it mentioned that AVC-501 may not consider dropped frames in the simulation. This is indeed the case in our simplified simulation, but in our prototype, dropped frames are common practices of real coders and the combiner can handle it through the frame synchronization mapping scheme we proposed in AVC-560.

5. In TD-43, sec. 6.5, it quoted AVC-501 as stating that synchronous terminal clocks ensure the same frame rates for different QCIFs. This is again only a simplified assumption. In the prototype, physical layer synchronization and timing recovery is handled by the network (either Ethernet or ISDN) interface and is transparent to software implementation. The combiner is "synchronous" to the users' video coders in a higher layer sense.

6. In TD-43, sec. 7, it concerned that the QCIF video rate of 17.2 kbps suitable to support an ISDN BRI operation is not sufficient for videoconferencing purposes. This is true but the situation applies to both transcoding and combining. According to our simulations, a CIF sequence consisting of 4 QCIF sub-sequences and coded at 4 times the rate does not offer better quality. The solution is to use a higher rate, such as 32/128 or 64/256 kbps. We have found 64/256 kbps operation to give quite reasonable picture quality in our experiments. If higher-rate isochronous packet network is available, the rate distribution will not be a problem. If ISDN is used, PRI rates of 6B or higher is more suitable to support multi-point video.

IV. Delay, Quality and Cost

From our experiments, we agree that the delay and quality differences between a combiner and a transcoder is not too much. Whether the differences may cause any service artifact remain to be investigated.

The cost or complexity issue, however, strongly favors the combiner approach. A transcoder requires 4 sets of (QCIF or CIF) decoders, a multiplexer, and a CIF encoder. while a combiner does not require any encoder or decoder, and can be implemented totally in software. Our prototype uses a simple PC, which operates in a primitive DOS environment. But it can still handle video combining with a comfortable processing margin. We believe for the time being, there is no simpler and cheaper multi-point video solution than the combiner.

V. Coded-Domain Combining or Transcoding

The coded-domain combining can be implemented without changing the existing H.320 standard. Because of its low cost and simplicity, it is possible to establish and start the multipoint videoconferencing service sooner in the near future. We have demonstrated the operation of the coded-domain combining through a real-time experimental research prototype implemented using a simple PC.

Transcoding, however, has a higher combining flexibility and may become a better solution in the further future. Currently, there are few reported transcoding prototypes or products. How to make transcoding more viable in terms of lower delay, higher quality and less cost as compared to coded-domain combiner remains to be investigated.

Bellcore is working actively in both areas to promote multipoint video conferencing services as soon and as viable as possible. At this time, since both coded-domain combining and transcoding can be implemented without changing the existing H.320 standard and each has its own advantages, both can co-exist. Indeed, it is also possible for an MCU to have a combination of combiners and transcoders to provide better combined picture quality and less terminal capability constraints.

The main purpose of this contribution is to present to the video expert group the simplicity and service implications of a combiner. We suggest the experts group for ATM video coding not to act on recommending which one to use for multipoint videoconferencing now but let the market decide which one prevails and at what time frame.