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

Document AVC-500

May 1993

INTERNATIONAL TELECOMMUNICATION UNION
TELECOMMUNICATION STANDARDIZATION SECTOR
STUDY PERIOD 1993-1996

COM 15-
May 1993
Original: English

Question: 3/15

STUDY GROUP 15 - CONTRIBUTION

SOURCE: BELLCORE (PROPOSED USA)
TITLE: MULTIPOINT 4 QCIF TO CIF VIDEO MIXING

ABSTRACT

This contribution proposes a simple scheme for providing 2 x 2 split-screen continuous presence videoconferencing, and preliminary text for the next revision of H.230 and H.243. This draft reflects comments that we have accepted from a preliminary review through T1A1 (from AT&T, VideoServer, and VTEL). This contribution will be put through the US white contribution review process for the September Study Group 15 meeting. We are submitting this draft to the rapporteur's meeting to stimulate timely consideration.

A. DISCUSSION

First generation H.231/H.243 MCUs are video switching MCUs that do not process the video bit stream. Based on audio signal levels and other switching controls, the MCU selects and transmits one received video signal to each terminal. This limits users to viewing one location at a time. Furthermore, some users may dislike the service aspects of video switching (e.g. the picture freeze/update process, imperfect audio switching, and need for chair control).

An alternative to video switching is video mixing, where the MCU transmits a split-screen picture consisting of several received video signals to each terminal. This mode of operation, also called continuous presence, allows users to see several locations simultaneously (each with less resolution) and reduces the need for video switching. If the number of terminals in a conference is less than or equal to one plus the number of video signals mixed per terminal, then all locations can see all other locations without video switching. Continuous presence has been recognized as a desirable alternative mode of operation, but has not been widely implemented because of complexity, picture quality, and delay issues.

For the 2 x 2 split-screen, however, a special case of H.261 video can be mixed with minimal complexity, quality degradation, and added delay. The basic concept is to spatially multiplex four QCIF pictures into one CIF picture by manipulating the picture header and group of block (GOB) header information, without actual video processing. This results in a CIF signal whose bit rate is nominally the sum of the bit rates of the four QCIF signals.

Within the H.261 framework, if terminals can decode CIF signals and transmit QCIF signals at 25% of the video channel rate by filling 75% of the video channel, then the MCU can construct 2 x 2 split-screen

pictures with relative ease. The filling operation can be done within the H.261 error correction frame structure by transmitting error correction fill frames.

Note that we considered requiring terminals to transmit 3 error correction fill frames for each coded data frame, but rejected it because lower delay is possible without this constraint. That is, terminals should code at a quarter of the video channel rate, but transmit using the full video channel when there is video data to transmit. All video data for each frame should be transmitted without interruption (using MBA stuffing if necessary), and should not be held until data for the next frame is available.

The MCU will clearly have to terminate and regenerate error correction framing, buffer and synchronize received video frames, and handle some video controls differently (e.g. freeze and update). It may need to transmit blank QCIF positions in a 4-QCIF mix, but should not exceed the maximum frame rate that terminals can decode. We believe an asymmetric video rate constraint on terminals is a necessary condition for this to work. We need to understand implementation options in more detail to determine if additional constraints or controls are necessary. Section B proposes a service flow in terms of draft H.230/H.243 text. Section C identifies issues for further study.

B. PRELIMINARY PROPOSAL

We propose the following additions and changes to H.230:

- Asymmetric video capability (Asym-V-cap): Declared by terminals capable of transmitting QCIF using up to 25% of the video channel rate. That is, the number of video bits (including MBA stuffing) transmitted over any interval of X seconds (e.g. 1 sec) should not exceed (25 - epsilon)% of the total available video bits. Asym-V-cap should be ignored if CIF capability is not declared. Asym-V-cap is also declared to other MCUs by MCUs that are capable of operating in 4-QCIF cascade mode.

- Asym-V-On: Sent by MCUs to terminals that have declared Asym-V-cap to tell the terminal to transmit asymmetric video. Also sent by MCUs to other MCUs that have declared Asym-V-cap to tell the remote MCU to receive in 4-QCIF cascade mode.

- Asym-V-Off: Negate the Asym-V-On command.

- An MBE version of VIN that allows for multiple terminal numbers and picture positions, {start-MBE/N/<vin>/<M><T><P>...}, where <P> = position 1 to 4. This is only valid in 4-QCIF mode.

- MIP, Multipoint Indicate Priority: sent by an MCU to another MCU during 4-QCIF cascade mode to indicate the priorities of the four QCIF pictures (see proposed H.243 text).

We propose the following new section for H.243:

4.3 Video mixing

Video mixing in general is not precluded by this recommendation.

Suppliers may implement video mixing schemes that, when in use, may conflict with H.243 controls. However, the required MCV command, and the VCB command required under chair control, should stop video mixing and broadcast the selected video, if these commands are supported during video mixing. Additional controls that enhance video mixing operations may be recommended in the future. The following sections describe one special case of video mixing that is of particular interest because it offers a good quality vs. complexity tradeoff.

4.3.1 4-QCIF mode

This mode of video mixing relies on the terminal capability to transmit a quarter-rate QCIF signal while receiving a full-rate CIF signal (Asym-V-cap). From the default audio-switched mode, an MCU capable of 4-QCIF video mixing may switch into 4-QCIF mode:

a) After the number of asymmetric video capable terminals is within some preset range, and/or
b) After some administrative conditions have been met (e.g. a reservation parameter has been set).

The MCU should send Asym-V-On to each asymmetric video capable terminal and transmit a 4-QCIF mix to that terminal (Figure 1). It may transmit a single 4-QCIF mix to all of these terminals, or tailor the mix for each terminal (e.g. so that users don't see themselves). The MCU should also transmit a 4-QCIF mix to each CIF capable terminal that is not capable of asymmetric video.

The MCU should, based on audio channel activity, select and transmit a quarter-rate QCIF signal, or a full-rate QCIF signal, to QCIF only terminals. The full-rate video signals transmitted by terminals that are not capable of asymmetric video should not be mixed unless the MCU can optionally transcode full-rate signals into quarter-rate QCIF signals.

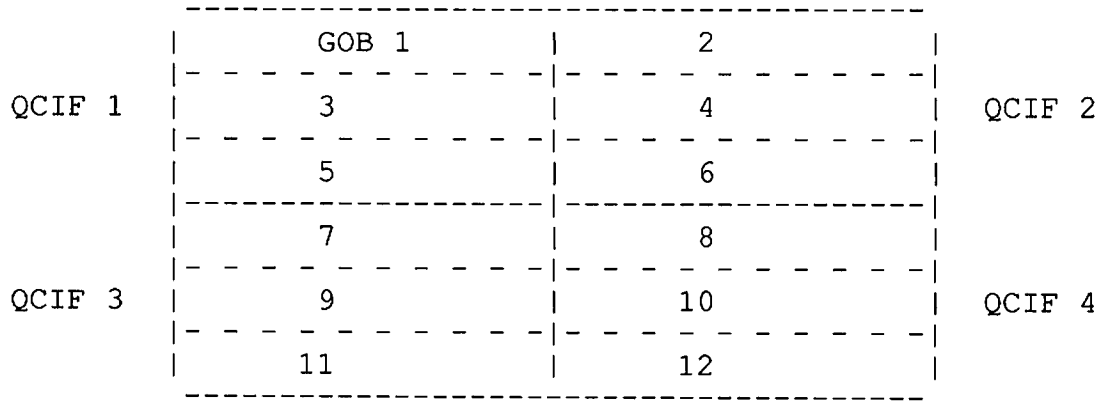

Figure 1/H.243
Four QCIF pictures in a CIF picture

If the number of asymmetric video terminals exceeds five, so that one location cannot see all other locations, the MCU may use any means (e.g. audio channel activity) to select the quarter-rate signals that are transmitted. If the number of asymmetric video terminals is less than four, the MCU should generate a neutral background for each missing QCIF position.

While transmitting a 4-QCIF mix, the MCU is not expected to send Freeze (VCF) or Freeze Release. The MCU may freeze each QCIF picture independently by stop sending video data for that picture. When the MCU selects a new quarter-rate video source for a 4-QCIF mix, it should send an Update (VCU) to that video source (or after an internal buffer overflow causing a video signal break). Upon receiving a VCU directed at a 4-QCIF mix, the MCU should send a VCU to each quarter-rate video source that makes up the 4-QCIF mix.

An MCU should not send VIN while transmitting a 4-QCIF mix. To indicate the terminal numbers of the individual QCIF pictures to an asymmetric video terminal (i.e. not to CIF terminals), the MCU should send an MBE version of VIN, {start-MBE/N/<vin>/<M><T><P>...}, where <P> = position 1 to 4.

4.3.2 Interaction with video selection

Interaction between video mixing and video selection in general is for further study. For the 4-QCIF mode, the MCU should observe the following guidelines.

4.3.2.1 Interaction with existing commands

1) MCUs may not support MCV, VCB, or VCS when operating in 4-QCIF mode.

2) For MCUs that support MCV or VCB when operating in 4-QCIF mode, MCV or VCB should override asymmetric video. That is, MCV and VCB should provide the broadcast service that they currently provide. For example, if a Terminal A operating in asymmetric video mode is selected as the broadcast video source, the MCU should send Asym-V-Off to this terminal before sending a VCU as part of normal video switching. The MCU should broadcast the full-rate signal from Terminal A to all other terminals. It may continue to send a 4-QCIF mix to Terminal A. Cancel-MCV or Cancel-VCB should restore 4-QCIF operation.

3) For MCUs that support VCS when operating in 4-QCIF mode, the service provided depends on the type of signal that the requesting terminal is receiving (4-QCIF mix or not 4-QCIF mix), and the type of the selected signal (quarter-rate or full-rate). The MCU may transmit one of the following to the requesting terminal:

   a) A replacement QCIF picture in a 4-QCIF mix (position undefined),
   b) A quarter-rate QCIF signal to a terminal not capable of Asym-V, or
   c) A full-rate signal from a terminal not capable of Asym-V.

4.3.2.2 4-QCIF mode video selection features [for discussion]

1) Chair control of multiple broadcast sources during 4-QCIF mode is a potential new feature. The chair may be able to broadcast terminal <M><T> at position <P> by sending an MBE version of the VCB command {start-MBE/N/<vcb>/<M><T><P>...}.

2) Terminal selection of multiple sources during 4-QCIF mode is a potential new feature. Terminals may be able to select terminal <M><T> at position <P> by sending an MBE version of the VCS command {start-MBE/N/<vcs>/<M><T><P>...}.

Interaction between <vcb>, <vcs>, and VCS is for further study.

4.3.3 Interaction with cascading

Interaction between video mixing and cascading in general is for further study. For the 4-QCIF mode, the MCU should observe the following guidelines:

1) Cascaded MCUs may not support 4-QCIF mode.

2) Cascaded MCUs may support 4-QCIF mode for locally connected terminals only and treat MCUs as terminals not capable of asymmetric video.

3) [For discussion] MCUs may support 4-QCIF operation in two-MCU cascades if both MCUs support the 4-QCIF cascade mode described as follows ("Asym-V" codes are used differently from the way they are used with terminals):

   a) MCUs that are capable of operating in 4-QCIF cascade mode should declare Asym-V-cap to other MCUs.

   b) After having received Asym-V-cap from another MCU, an MCU capable of 4-QCIF cascade mode may send Asym-V-On to the other MCU and transmit in 4-QCIF cascade mode. If a master MCU has been assigned, the slave MCU should follow the master.

   c) When transmitting in 4-QCIF cascade mode, an MCU may transmit up to four QCIF pictures in a 4-QCIF mix. For each inactive QCIF position, the MCU should skip the corresponding GOB headers (active QCIF positions may contain GOB headers with no video data).

      A priority designation of the four positions is necessary. For example, when the total number of asymmetric video terminals exceeds five and audio switching is used to select four QCIF pictures, the QCIF picture of the "current speaker" at the local MCU should be designated high priority and should be mixed by the remote MCU.

      One possible scheme is for the MCU to repetitively send MIP<N> to indicate the priorities of the four QCIF positions. BAS bits 4, 5, 6, and 7 of <N> indicate the priorities of QCIF positions 1, 2, 3, and 4, respectively. "1" means high priority, "0" means low. A 4-level priority scheme which allows for priority ranking is also possible but the exact coding and usage need further study. A triple symbol may be needed because of the limited range of SBE numbers.

   d) In the previous section, MCUs operating independently may generate different 4-QCIF mixes which may confuse users. If a single 4-QCIF mix is desired (at the risk of seeing one's self), and if a master MCU has been assigned, the master should designate all four positions of its 4-QCIF mix to have high priority. The slave should yield to the master, and switch the master 4-QCIF mix to all terminals.

      In the slave to master direction, the slave should designate the position of the "current speaker" to have high priority. The master may also send <vcs> to the slave to obtain a specific 4-QCIF mix. The master may use "<M><0><P>" to blank out position P coming from MCU M. The slave may transmit other available

quarter-rate signals in positions unspecified by <vcs>. This scheme should work for more than two MCUs.

e) MCUs should send Asym-V-Off to another MCU to stop transmitting in 4-QCIF cascade mode. This is necessary, for example, when a broadcast video source is selected by MCV or VCB.

C. ISSUES

The following issues have been identified:

1) This contribution makes the simplifying assumption that each terminal sends one QCIF picture (one picture and one format minimize signaling, QCIF implies bounded motion vectors). Operational constraints are placed on rate control, which means no change to H.261. This assumption can be relaxed in the following ways:

- Allow terminals to send more than one picture, each with bounded motion vectors (e.g. from multiple cameras) and video rate proportional to picture size. This implies a single terminal number <M><T> may not be enough for video source selection.
- Allow pictures to contain different subsets of "active" GOBs in a CIF picture, each with bounded motion vectors and video rate proportional to picture size. This requires signaling to identify active GOBs and determine a common video mixing mode.

This contribution describes what we believe to be the simplest useful video mixing service. However, we recognize the desire for more flexible services. A decision needs to be made on a target service which accounts for increasing complexity and changes to existing standards. If a more flexible service, or an hierarchical tier of services, is desired, contributions on more flexible control schemes are solicited.

2) A related service issue for further study is the interaction between video mixing and the H.261 split-screen indicator bit, for which we have no procedures.

H.261 currently indicates the transmitted picture type (e.g. split-screen, CIF/QCIF) in PTYPE. It may be desirable to define an additional picture type for a quarter-rate QCIF signal, or for a more general sub-CIF signal with bounded motion vectors and proportional video rate. This requires changing H.261.

3) Although the proposed video mixing MCU does not have to transcode video, it does have to deal with variable length and generally asynchronous incoming video frames. The implementation details need to be better understood. For example, one key implementation issue is deciding when to transmit a large intraframe. An MCU may transmit blanks until the frame is fully received, and possibly incur longer delay, or start transmitting earlier (e.g. after receiving one or two GOBs) and buffer the other inputs. Note that once an MCU starts to transmit a QCIF frame, it should not run out of data because the terminal is also transmitting at the full video channel rate (assuming the terminal only transmits one sub-CIF picture).

As we understand the implementation issues better, we may consider changes to H.261. One possible change is to reduce the maximum QCIF picture size from 64 Kbits to, for example, 16 Kbits. Another possible change is to relax the H.261 requirement to transmit all GOBs in sequence. We hesitate to propose any change to the hypothetical reference decoder because we do not believe it is a good model of the video mixing operation.

Additional BAS controls are also possible. Currently, the only feedback command that an MCU can send to a terminal is VCU. The problem with VCU is that the terminal will most likely respond with a large intraframe, which does not help the MCU to reduce its buffer content. It may be desirable to introduce an SBE skip frame command that tells the terminal to use interframe coding, but temporarily slow down the frame rate. This may give the MCU more effective feedback control.

End