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

Doc. #240
May 1987
Original: English

Question: XV/4 Specialists Group

STUDY GROUP XV - CONTRIBUTION

Source: United States of America

Title: Proposed framework for m x 56/64 kbps videophone standard

1 INTRODUCTION

The standardization work on m x 56/64 kbps can be structured into three parts:

1. Communication procedure and frame structure.

2. Hardware related parameters, such as picture format.

3. Complete specification of standard algorithm.

The goal should be to agree on the first two items within the current study period, such that manufacturers can develop products to the standard as soon as possible. A complete algorithm specification should be finalized during the next study period (1989-92), possibly through an accelerated procedure by 1990.

It is of prime importance to agree on parameters that allow inexpensive implementations for both NTSC and PAL/SECAM equipment, such that the CCITT standard will also become the de facto standard.

2 APPLICATIONS

The applications for m x 56/64 kbps video coding can be divided into videophone and videoconferencing. The videophone application is characterized by head-and-shoulders pictures shown on a monitor that is typically smaller than 12 inches. The videoconferencing application has the need to transmit scenes containing small groups of people, typically up to three people seated side by side, or six people with a split-screen arrangement.

Both applications have the need for graphics capabilities.

A limited graphics capability is provided by the moving video, but most applications also have the need for high resolution graphics that can transmit a typewritten page.

It is also important to be able to connect external equipment to the system via data ports. The external equipment can be a facsimile machine, a data terminal, or some other equipment.

3 BIT RATE

In public networks, the bit rates of primary interest are m x 56/64 kbps, where m=1 or m=2. In private networks, higher values of m are also of interest.

The main focus is to study integrated video, audio, and data at rates between 56 and 128 kbps. Assuming an audio bit rate of 16 kbps for m=1 and between 16 and 64 kbps for m=2, the video coding algorithm should be able to operate from 40 to 112 kbps.

4 PERFORMANCE GOALS

The video resolution should be sufficient to handle a conference scene with three people seated side by side.

The goal for the frame rate should be 10-15 frames per second. Among other things, this will allow lip synchronization to be maintained.

A maximum one-way processing delay of 250 ms should be strived for.

5 COMMUNICATION PROCEDURE

A communication procedure makes it possible to negotiate parameters between the communicating terminals. By knowing the capabilities of the other party, a set of parameters for the session can be agreed upon. The procedure should also involve a decision on the algorithm to be used. This makes it possible to upgrade the standard as the state of the art is advanced.

It also makes it possible to use proprietary algorithms if both parties are suitably equipped; the standard algorithm can be used as a fallback.

6 FRAME STRUCTURE

The frame structure should fulfill the following requirements:

1. Allow dynamic bit allocation for video, audio, and data.

2. Procedure to synchronize two transmission channels. This will allow Type 3 applications without requiring Time Slot Integrity to be maintained in the network.

3. Ensure correct operation in presence of transmission errors.

4. Byte oriented format to allow easy microprocessor implementation.

5. Permit forward compatibility with emerging ISDN recommendations for terminal parameter negotiating sequences and for dynamic network allocation control.

The above requirements may be satisfied in several ways among which would be a packet format. Since efficient coding algorithms require virtually error free transmission, an error correction procedure should also be part of the frame structure.

7 PICTURE FORMAT

The picture format should be chosen based on the particular applications and requirements foreseen for the m x 56/64 kbps videophone service. A new Common Intermediate Format can be defined (CIF-64), or a dual mode approach can be used with different formats for NTSC and PAL/SECAM.

A Common Intermediate Format implies standards conversion both in transmitter and receiver independent of TV standards. The dual mode approach typically involves a conversion in the receiver if the transmitter and receiver use different TV standards.

A new Common Intermediate Format should be selected based on the following considerations:

1. Resolution sufficient for small videoconferences with up to three people seated side by side.

2. Economical implementation for NTSC and PAL/SECAM.

3. Simple relationship with the digital studio standard (CCIR 601); although this might not be important for codecs with integrated cameras and monitors.

Commercial 56 kbps codecs and the CCITT 2 Mbps Part I standard (H.120) indicate that the sampling frequency should be around 5 MHz to achieve the resolution goal. This resolution also allows a limited graphics capability using the moving video.

Simple implementation for both NTSC and PAL/SECAM implies that the vertical resolution should be kept at or below 240 lines, such that only one field needs to be accessed in both standards. If the number of lines is chosen as a multiple of 48, the format will have simple relationships with both NTSC (240 lines) and PAL/SECAM (288 lines). 5 MHz horizontally corresponds to 260 pixels per line. An aspect ratio of 4:3 implies 195 lines vertically to achieve the same resolution in both dimensions.

Simple relationship with CCIR 601 suggests a sampling frequency of 4.5 MHz (1/3 x 13.5) or 5.06 MHz (3/8 x 13.5).

Two formats seem particularly attractive:

270 pixels x 240 lines (Could be cropped to 256x240 which has become a defacto standard for codecs in North America):
Horizontal sampling frequency 3/8 x 13.5 MHz = 5.0625 MHz.
Vertical conversion ratio for NTSC: 1:1
Vertical conversion ratio for PAL/SECAM: 6:5

240 pixels x 192 lines:
Horizontal sampling frequency 1/3 x 13.5 MHz = 4.5 MHz.
Vertical conversion ratio for NTSC: 5:4
Vertical conversion ratio for PAL/SECAM: 3:2

Other formats that have been discussed within the CCITT Specialists Group are 180x144 and 360x288. The former is useful for face-to-face applications (m=1), but does not give adequate resolution for groups of people.

The 360x288 format has a higher resolution than required by the foreseen applications. This would make the codec equipment more expensive to manufacture, since the cost of the signal processing part of the codec is roughly proportional to the number of pixels per second. It would also make it more difficult to reach the desired frame rate goal of 10-15 frames per second.

For NTSC equipment an additional cost would be incurred, since both fields need to be captured and processed to obtain the 360x288 format. The processing would need to be motion adaptive to avoid artifacts due to the time difference between the two fields.

A chrominance subsampling of 4:1 horizontally and 2:1 or 4:1 vertically seems possible without significant degradation.

For graphics operation, it is proposed that a separate still

frame mode is used that would double the horizontal and vertical resolutions. Interactive pointing at documents would be handled by means of superimposed cursor symbols.