CCITT
STUDY GROUP XV WP XV/1
Experts Group for ATM Video Coding
London 28-30 October 1992

Document # AVC-391

English Version

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
ORGANIZATION INTERNATIONALE DE NORMALIZATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND ASSOCIATED AUDIO

ISO/IEC JTC1/SC29/WG11
MPEG92/
October 1992

SOURCE: AT&T Bell Laboratories
PURPOSE: Informational
TITLE: PUBLIC SWITCHED TELEPHONE NETWORK (PSTN) VIDEOPHONE

1. Introduction

The contribution addresses the growing need for standardized videotelephony service over the analog public switched telephone network (PSTN), including the ability to manage the telephone line data bandwidth to allow the incorporation of other data services in a generic process. We suggest a modulated data scheme which incorporates video, audio, and data to be sent digitally, using video compression, audio compression, and application-specific data streams. It is our strong belief that such a system can be realized in a cost-effective manner appropriate for the consumer market.

Fundamentally, the PSTN characteristics set a limit to the total amount of data that can be transported through the network. The analog loop between the central office and the telephone equipment often limits the performance, though in international calling, the long-distance network will affect the performance because of the prevalent use of ADPCM. We will discuss how our proposal addresses ADPCM in the section on modem algorithm.

Because the available data bandwidth is limited, the video is constrained by two variables, as perceived by the user. First is the image quality and second is the image frame rate. These two variables are inversely proportional; that is, a sharper image requires more data and hence has a lower frame rate, while a fuzzy image requires less data and hence has a higher frame rate. We fundamentally believe that the user should be able to choose how to view the far end party - with more clarity or more motion or a setting in between the extremes.

Further, the user should be able to decide if other data applications are desired, which in turn will affect the bandwidth available to the video. Currently, we propose a constant rate audio coder because of quality concern, but in the future, the audio bandwidth can be negotiated so that the user optimizes the bandwidth allocation according to the situation. We feel dynamic bandwidth allocation will provide the best audio and video and data performance for the user.

2. Video Algorithm

We propose that the video coder should be a constant-quality image coder, rather than a constant-frame rate coder. This allows the user to make an intuitive tradeoff between image quality and image motion.

We would like to see the equivalent intermediate format based upon cost-effective components and formats that are simple to use domestically and internationally. This means that the intermediate format should be easily generated from the various television standards in use throughout the world. An important consideration in the intermediate format is that since the spatial resolutions are low, defects which would be more apparent at higher processing resolutions will not be perceived as contributing to coding defects; i.e., coding artifacts are dominated by the lower processing resolution.

2.1 Image Coding Features

The basic video compression technique uses 8 x 8 discrete cosine transform (DCT), motion estimation, quantization, normalization, and lossless compression (Huffman encoding). The attached figure gives a block diagram of the process.

The luminance input data is a matrix of 128 pixels by 112 lines.

The chrominance input data is a matrix of 32 pixels by 28 lines, for each.

2.2 Motion Estimation

Motion estimation is based on sub-pixel resolution. The best match is found, and if acceptable, a motion vector is sent. The motion estimator is based on a 16 x 16 pixel area scanned 6 pixels in any direction. Motion estimation is not required of the encoder, but the frame rate will be much lower without it.

2.3 Video Frame Structure

Each video frame transmission is bounded by a Frame Sync Word (FSW). The importance of the FSW is that it forms a layer of error control and recovery, and prevents the user from seeing a frame which does not conform to the expected format. The process also allows for the FSW to be sent at anytime, thus allowing for robust and automatic video refresh and error recovery without having the user see a image with defects. I.e., until a video frame is detected as having been received complete and correct, the output that the user sees will not be updated. This does not address errors within a frame that might pass for a valid frame of video data - this is covered in the modem and protocol sections.

Inside the video data frame, information related to motion vectors, DCT coefficients, and block refresh are sent after zig-zag, quantization, normalization, and Huffman encoding. The process is reversed at the receiver.

3. Audio Algorithm

It is important to have good quality speech in videotelephony for all types of speakers; i.e. toll-grade quality for male, female, and child voices.

3.1 CELP +

The CELP+ (Code Excited Linear Prediction Plus) meets toll-grade quality requirement for men, women, and children speakers. CELP+ is based on a GSM digital cellular submission. It works at 6800 bps.

3.2 Audio Frame Structure

The basic audio frame compresses 20 ms of 8 KHz sampled speech into 17 filter coefficients, each one byte long.

4. Modem Algorithm

The current modem uses a multi-dimensional coding at 2800 baud and 3200 baud to give 16.8 and 19.2 kbps, respectively. The modems are full-duplex, synchronous type. In cases where 3200 baud is not possible, the videophones fall back to 2800 baud. In testing over international connections using data optimized ADPCM equipment, the 2800 baud has been shown to operate as well as V.32bis modulation of 2400 baud at 14.4 kbps. For international calling, data optimized ADPCM equipment is needed for either 14.4 or 9.6 kbps using V.32bis modulation, or for the 2800 baud multidimensional coding. This is the same network equipment that makes Group 3 facsimile international-capable at 9600 bps.

4.1 Video Telephony Applications

We feel that analog videotelephony places different constraints on the modem use than traditional data modem. Essentially, the system is now a real-time system where a fast start-up process is desired to minimize the amount of time the user waits to begin the videotelephony session. A traditional start-up has a more complete line characterization process, including a phase-lock loop for very slow phase roll. The traditional start-up would keep the users waiting too long before they can start to talk and see each other. The reduced start-up will lessen the successful connection rate by a small fraction because the long-distance networks are now dominated by digital carriers. The source of modem impairments is now more concentrated in the local loop where phase roll is a small problem; instead, PCM noise, digital slip, line coil loading, and site-specific impairments are dominating effects.

4.2 Emerging Modulation Techniques

Currently, two proposals are under study by CCITT for the basis of V.Fast. It is our intention to incorporate V.Fast into our product when it becomes a standard. We are motivated to incorporate V.Fast because of our belief that V.Fast offers a flexible data bandwidth which is in accord with our system design outlook. However, we chose the current proprietary modulation because it offers benefit to the user without affecting the connectivity greatly.

To compensate for the local loop impairments, new techniques are being developed that will allow the modem to characterize the local loop and add pre-emphasis to both transmit and receive filters. These loop balancing techniques will improve connectivity and are independent of the modulation scheme, so we can incorporate these improvements without burdening the users with a long start-up sequence.

5. Communications Protocols

As stated above, we believe that the protocols need to provide flexibility for adding data services in addition to the video and audio. Additionally, the protocol needs to provide flexibility in allocation of the data bandwidth.

It is also known that the video data stream does not carry much redundancy between video data frames so that bit errors represent long-lived image distortions on the display (on the order of several seconds) which are likely to be objectionable. Likewise, the CELP audio coders are sensitive to bit errors and framing errors. However the audio latency is small so there is little effect carried from one frame to the next.

For the reasons stated above, we believe that a link access protocol (LAP) and some sort of network layer are required. The LAP assures that the data is passed onto the video codec without errors, hence without image distortions. The audio data can be accepted with errors, ignoring the short-term distortions. The purpose of the network layer is to add other application-specific data streams, such as end-to-end signaling, user data, data-base services, etc.

We have also defined end-to-end control and status signaling within the system for identifying the capabilities within each set. This allows for backward compatibility and the means to negotiate using technology advances as they become available.

5.1 Link Access Protocol

The current definition is LAP-B. We are examining LAP-D and extensions to it that optimize it for videotelephony. Improvements in transport, such as forward error correction can be added here in a consistent manner.

5.2 Network Access Protocol

The current definition is X.25. In looking at LAP-D, we are considering adding a network management function to LAP-D.

6. Data Transport Services

We believe that the ability to add new services along with videotelephony is of great interest. The protocols we are considering allow for a systematic approach to integrating new services while maintaining compatibility with equipment already in service. While older equipment may not be able to take advantage of such services, we preserve the function for which they were bought.

Because the data transport is realized in virtual circuits, the dynamic allocation of data bandwidth can be managed completely at the transmitter. Through the use of the end-to-end control channel, requests for different allocations can be received from the far-end. This system provides the user with the ability to direct the bandwidth allocation as best fits the situation. For example, the user can devote all the bandwidth to video to get the best motion rendition, or may choose to suspend video temporarily to get higher alphanumeric throughput.

Additionally, we can define types for each circuit. For example, we can define a video circuit, and within that definition we can differentiate between video, JPEG, and other types of image data streams.

7. Summary of Important System Features

- We believe in a constant-quality coder rather than a constant-frame rate coder, and that the user should choose the tradeoff between image quality and frame rate.
- Toll-grade voice compression for all speaker types.
- Robust data transmission with flexibility.

  The use of framing structures over the PSTN is questionable. We have experience with a similar structure and can comment as follows: under noise-free conditions, there is not a problem. However, we find the real-world connections often have noise. Missing frames will cause video and audio distortions which are objectionable to the user. The recovery from frame mis-alignment is not straightforward, as is becoming clear with other bit-oriented multiplexed protocols, because the detection of frame alignment loss can be difficult, causing the data streams to become intertwined. While it may appear that data latency will be less for a multiplexed protocol, to compensate for framing loss and other error recovery mechanisms, the receiver must add delay to buffer input data until it is determined that received data is good. On the whole, the latency will be equivalent to a more traditional link access protocol, but without the robustness and recovery mechanisms.

  Finally, we believe that the future of videotelephony over the analog PSTN will involve the ability to add in new services, which will not be easy with British Telecom implementation.