CCITT SG XV

Doc #82

Specialists Group on Coding for Visual Telephony

Source: UK, FRG, France, Italy, Netherlands, Sweden

Title: **Specification for the Flexible Prototype 2nd Generation Videoconference Codec**

---

This document presents the state of the discussions between the laboratories in Europe who are proposing to construct hardware codecs and contribute to the work of the specialists group.

The aim has been to produce a specification which essentially is the minimum implementation with which all laboratories will comply. The specification will provide considerable scope for optimisation and experimentation. It can be seen that the document is not yet complete but is being presented because it is proposed that the CCITT group use the same approach to its work and adopt this document, or a similar one, as the working draft specification for the hardware.

This concept of a "minimum" codec worked well during the development of the H110/120 codec.

We would expect that the final specification to be produced in early 1988 would be significantly different as it would reflect advances made during 1986/87. It would also have the areas of flexibility removed.

CONTENTS


1 Source Coder
2 Video Multiplex Coder
3 Transmission Coder


# 1 SOURCE CODER

## 1.1 Source format.

1.1.1 The format to be coded is 288 lines, 30000/1001 (approximately 29.97) non-interlaced pictures per second - the Common Intermediate Format. The tolerance on the picture frequency is +/- 50 ppm.

1.1.2 Pictures are coded in component form, these components being luminance (Y) and two colour difference signals (R-Y & B-Y). There are 360 luminance pels per line with an orthogonal sampling pattern. The colour difference sampling parameters are 180 samples per line, 144 lines, orthogonal. Both R-Y and B-Y samples are sited such that their block boundaries coincide with luminance block boundaries.

1.1.3 Source format codes correspond to CCIR 601:

Black = 16
White = 235
Zero colour difference = 128
Peak colour difference = 16 and 240

Codes outside the above ranges will be accepted but may be modified by the coder to avoid emulation of reserved codewords.

## 1.2 Video Coding Algorithm.

1.2.1 The video coder algorithm is shown in a generalised form in figure no. 1. The coding scheme utilizes a block transform. The predictor may incorporate movement compensation.

1.2.2 Motion compensation is optional at the coder. The decoder will accept one displacement vector for each luminance block of size m pels by n lines. The maximum displacement vector is +/- 16 pels and +/- 16 lines. Only integer values of the horizontal and vertical components of the vector are permitted. (Note: The encoding method for transmission of displacement vectors may restrict the vectors to a subset of these 1089 possible values that the decoder hardware can accomodate. This will be specified later.)

Chrominance displacement vectors?
(None, independent or derived from luminance.)

A positive value of the horizontal component of the displacement vector signifies that the prediction is formed from pels in the previous picture which are spatially to the right of the pels being predicted.

A positive value of the vertical component of the displacement vector signifies that the prediction is formed from pels in the previous picture which are spatially below the pels being predicted.

### 1.2.3 Coder

After prediction, the resulting difference picture or the original picture is subdivided into blocks which are classified as transmitted or non-transmitted. The classification criterion is not part of the specification, being left free to equipment designers and may be varied dynamically as part of the data rate control strategy. Transmitted blocks are coded by a transform based scheme. Coding block size for luminance is i by j sample values.

Chrominance transform blocks contain the same number of pels as luminance ones and so cover four times as much picture area.

### 1.2.4 Transformer

The transformer shall be implemented in a flexible manner so that a number of different transforms, or configurations of a particular transform in terms of bits per stage, can be investigated. All hardware should be equivalent in performance to classical matrix multiplication, so fast algorithms are not excluded provided they can be constructed to be as flexible.

For the purposes of compatibility it is only necessary to specify the inverse transform. As a preliminary specification the hardware should allow for 12 bit resolution in (ie coefficients) and 9 bits out. (This is subject to verification that it does not imply a greater accuracy than 16 by 16 bit multiplications.)

### 1.2.5 Quantizer

The number of inverse quantisers provided in the decoder shall be 32. Each quantiser has 12 bits input and up to 16 bits out.
(The same or different for luminance/chrominance?
Adaptive/non-adaptive)

## 1.3 Data rate control and subsampling modes.

The exact method of assessing the encoder data generation rate need not be specified but the specified minimum size of decoder buffer must incorporate an allowance for latency in the assessment and control loop. Hence, any requirement to constrain overall system delay may effectively preclude some schemes.

Control information is carried by side information - not derived recursively from received data.

Horizontal subsampling - picture, group of blocks or block basis?
Vertical subsampling " " " " " " "
Temporal subsampling - picture basis only. Interpolated pictures are not placed in the picture memory.
Quantizer selection - see also 1.4.
Block significance criterion - not part of the specification.

The coding algorithm will automatically permit the full quality of the source format specified in section 1.1 to be realised on still pictures.

1.4 Forced Updating

There is no separate coding scheme for forced update. This function is achieved merely by forcing the use of the intra-picture mode of the coding algorithm. Since the decoder cannot distinguish between normal and forced update blocks, then there are no parameters to specify which are unique to forced update. The bitrate allotted to forced updating, the sequence in which blocks are updated etc, are not specified.

In order that the quality of forced update blocks can be sufficiently high at all times, quantiser selection must not be dependent solely on buffer fill state.

## 2 VIDEO MULTIPLEX CODER

2.1 The video multiplex coder has the following tasks:

1. Block address coding
2. Video data formatting and serialising
3. Synchronisation - picture/line
4. Motion vector coding
   (Absolute, differential, run length etc)
5. Side channels for indicating dynamic coding parameters eg subsampling modes, quantizers, buffer state etc. (Permanent or transient channels? Transient channels require care in switched multipoint.)

2.2 Video Multiplex Arrangement

2.2.1 Picture start code (PSC)

Unique sequence of f bits [Buffer state] [Temporal ref.] [Type]

The buffer state is a 6 bit number representing the encoder buffer fullness in 1Kbit units at the beginning of this picture.

The temporal reference is a 3 bit number representing the time sequence, in Common Intermediate Format picture periods, of a particular picture.

Type is a Variable Length Code which allows block attributes to be applied to all blocks within a picture.

All PSCs are transmitted.

2.2.2 Group of Blocks Start Code (GBSC)

Unique sequence of g bits [Group Number] [Type]

A group of blocks consists of two lines of luminance blocks, one line of (R-Y) blocks and one line of (B-Y) blocks.

The group number is a x bit number representing the vertical spatial position, in units of groups, of the current group of blocks.

Type is a Variable Length Code which allows block attributes to be applied to all blocks within a group of blocks.

All GBSCs are transmitted.

2.2.3 [Block Address]

Block Address is a Variable Length Code indicating the relative position of a transmitted block relative to the previous transmitted block or the absolute position within the group of blocks if it is the first transmitted block in that group.

Addresses corresponding to absolute positions greater than ... are considered to be chrominance blocks.

2.2.4 Block Type

A Variable Length Code representing the type of the block.

1. Intra-picture coded block
2. Inter-picture coded block
3. Motion compensated block
4. Motion compensated with coded residue
5. Future expansion on
6. Future expansion off

The decoder shall be designed to ignore all data between types 5 and 6 and also between type 5 and the next GBSC.

2.2.5 Block Data

Data specifying motion vectors, scanning class, quantiser type and transform coefficients.

2.2.6 Block Delimiter Information

The means to identify the end of all data pertinent to one block. (This may be contained in the block data.)

2.3 Multipoint Considerations
(Fast updates etc)

## 3 TRANSMISSION CODER

3.1 The transmission coder assembles all data and interfaces to the digital line transmission system.

3.1.1 The data rate is m x 384 Kbit/s where m is an integer between 1 and 5, both inclusive.

3.1.2 The codec output clock rate source shall be switchable between either a free running internal source or a source synchronised to the received data from the network. The mechanism for this switching is ....? (Bit .. of timeslot ..?)

3.1.3 When in free running mode the tolerance on output clock rate will be +/-50 ppm of nominal.

3.1.4 When in synchronised mode the synchronism should be maintained when the frequency of the received data clock is within +/- 50 ppm of nominal.

3.2 Framing structure

As per CCITT Study Group XV WP XV/1 Doc #58R plus the following coding for the applications channel:

list of codec attributes/facilities/parameters needing transmission from transmitter to receiver. No return path is assumed. Currently this list includes only encryption.

Operability with audioconferencing. Timeslot positioning to CCITT Rec G737.

3.3 Video data buffering

The size of the transmission buffer at the transmitter is switchable from 8Kbits to 64 Kbits, both inclusive, in steps of 8Kbits. Buffer size should be related to the transmission rate (overall - not video) to ensure acceptable system delay.

3.4 Video clock justification
Not required for European version?

3.5 Optional full spatial resolution mode data for quasi-stationary pictures.

To be specified later if required. Transmission will be by data timeslot dynamically taken from transmission multiplex. This mode to be compatible with audioconference.

3.6 Audio

As per CCITT Draft Rec. G72X. Bandwidth 50 - 7000Hz
The audio timeslot can include data and signalling.

Flexible testbed hardware need only incorporate?
(64Kbit/s A-law
64Kbit/s u-law
56Kbit/s sub-band ADPCM according to CCITT Draft Rec G72X)

3.7 Error handling

Video coding strategy to be error resilient without internal or external error corrector. Note that demand refresh for error correction requires back channel.

3.8 Encryption

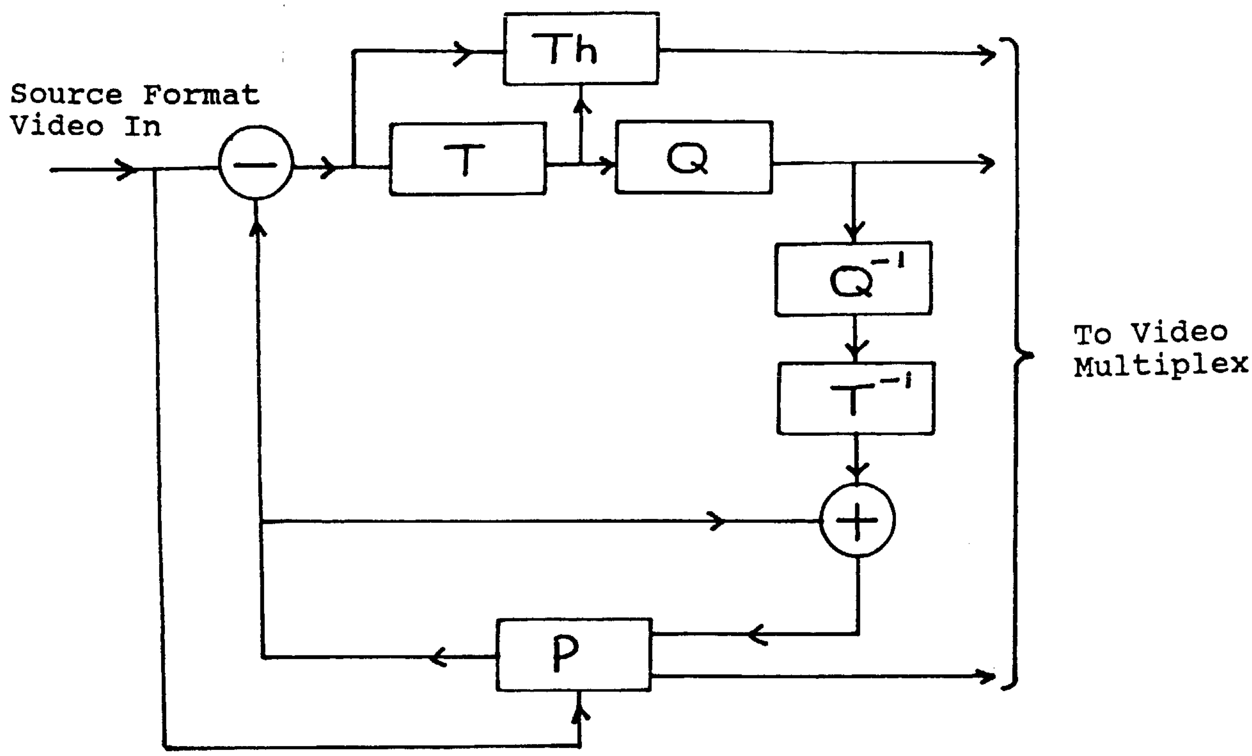
3.9 Data transmission

Framing structure to allow 2 data ports of 64Kbit/s each, though picture quality constraints may require only one to be available at 384kbit/s. TS17 is proposed in order to retain compatibility with audioconference at 2x64Kbit/s.

3.10 Network Interface

Access will be at 2Mbit/s with vacated timeslots. CCITT G703
(Access arrangements at 1.5Mbit/s ?)

END




| | |
|---|---|
| Th | Threshold |
| T | Transform |
| Q | Quantiser |
| $Q^{-1}$ | Inverse Quantiser |
| $T^{-1}$ | Inverse Transform |
| P | Predictor |

Figure 1.