CCITT SGXV
Working Party XV/1
Specialists Group on Coding
for Visual Telephony

Document #296
January 1988

Title: Second draft of H.12x

The second draft of H.12x (nx384kbit/s video codec) dated 11 January 1988 is attached. It has been prepared taking account of received comments on the first draft.

D G Morrison
BTRL
UK

Title: CODEC FOR AUDIOVISUAL SERVICES AT Nx384KBIT/s

[Note. Text enclosed in square brackets is not part of the recommendation. It is included as the drafter's comments, questions and explanations.]

[Changes from the first draft dated 23 November 1987 are indicated by vertical lines in the left hand margin.]

Contents

1. Scope
2. Brief Specification
3. Source Coder
4. Video Multiplex Coder
5. Video Data Buffering
6. Transmission coder

The CCITT,

considering

that there is significant customer demand for videoconference service;

that circuits to meet this demand can be provided by digital transmission using the H0 rate or its multiples up to the primary rate;

that ISDNs are likely to be available in some countries that provide a switched transmission service at the H0 rate;

that the existence of different digital hierarchies and different television standards in different parts of the world complicates the problems of specifying coding and transmission standards for international connections;

that videophone services are likely to appear using basic ISDN access and that some means of interconnection of videophone and videoconference terminals should be possible;

that the H.120 recommendation for videoconferencing using primary digital group transmission was the first in an evolving series of recommendations;

appreciating

that advances are being made in research and development of video coding and bit rate reduction techniques which will lead to further recommendations for videophone and videoconferencing at multiples of 64kbit/s during subsequent study periods, so that this may be considered as the second in the evolving series of recommendations;

and noting

that it is the basic objective of CCITT to recommend unique solutions for international connections;

recommends

that in addition to those codecs complying to recommendation H.120, codecs having signal processing and interface characteristics described below should be used for international videoconference connections.

Note: Codecs of this type are also suitable for some television services where full broadcast quality is not required.

1. Scope

This recommendation .....

2. Brief Specification

An outline block diagram of the codec is given in figure 1. [Based on fig.1 of #249, but with TSC and PrF combined into 'preprocessing'. Similarly on decoding side.]

2.1 Video input and output

To permit a single recommendation to cover use in and between 625 and 525 line regions, pictures are coded in one common intermediate format. The standards of the input and output television signals, which may, for example, be composite or component, analogue or digital and the methods of performing any necessary conversion to and from the intermediate coding format are not subject to recommendation.

2.2 Digital output and input

Digital access is at the primary rate of 1544 or 2048 kbit/s with vacated timeslots in accordance with I.431.

2.3 Sampling frequency

Pictures are sampled at an integer multiple of the video line rate. This sampling clock and the digital network clock are asynchronous.

2.4 Source coding algorithm

A hybrid of inter-picture prediction to utilise temporal redundancy and transform coding of the remaining signal to reduce spatial redundancy is adopted. The decoder has motion compensation capability, allowing optional incorporation of this technique in the coder.

2.5 Audio channel

Audio is coded to CCITT Rec. G722. This is combined with control and indication information and conveyed in one 64 kbit/s timeslot which conforms to CCITT Rec. H.221.

2.6 Data channels

H.221 permits part of the 64kbit/s timeslot carrying the audio to be used for auxiliary data transmission.

Additionally, one of the timeslots normally used for video may be reassigned as a 64kbit/s data channel. The possibility of further such channels is further study.

2.7 Symmetry of transmission

The codec may be used for bidirectional or unidirectional audiovisual communication.

2.8 Error handling

For further study.

2.9 Propagation delay

For further study.

2.10 Additional facilities

For further study.

3. Source Coder

3.1 Source format

The source coder operates on non-interlaced pictures occuring 30000/1001 (approximately 29.97) times per second. The tolerance on picture frequency is ±50 ppm.

Pictures are coded as luminance and two colour difference components (Y, CR and CB). These components and the codes representing their sampled values are as defined in CCIR Rec.601.

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

These values are nominal ones and the coding algorithm functions with input values of 0 through to 255.

For coding, the luminance sampling structure is 288 lines per picture, 352 pels per line in an orthogonal arrangement. Sampling of each of the two colour difference components is at 144 lines, 176 pels per line, orthogonal. Colour difference samples are sited such that their block boundaries coincide with luminance block boundaries as shown in figure 2. [fig.2 of #249] The picture area covered by these numbers of pels and lines has an aspect ratio of 4:3 and corresponds to the active portion of the local standard video input.

Note: The number of pels per line is compatible with sampling the active portions of the luminance and colour difference

signals from 525 or 625 line sources at 6.75 and 3.375 MHz respectively. These frequencies have a simple relationship to those in CCIR Rec.601.

3.2 Video source coding algorithm

The video coder algorithm is shown in generalised form in figure 3. [fig.4 of #249] The main elements are prediction, block transformation, quantisation and classification.

The prediction error (INTER mode) or the input picture (INTRA mode) is subdivided into 8 pel by 8 line blocks which are segmented as transmitted or non-transmitted. The criteria for choice of mode and transmitting a block are not subject to recommendation and may be varied dynamically as part of the data rate control strategy. Transmitted blocks are transformed and resulting coefficients quantised and variable length coded.

3.2.1 Prediction

The prediction is inter-picture and may be augmented by motion compensation (3.2.2) and a spatial filter (3.2.3).

3.2.2 Motion compensation

Motion compensation is optional in the encoder. The decoder will accept one vector for each luminance block of 8 pels by 8 lines. The range of permitted vectors is for further study.

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

Motion vectors are restricted such that all pels referenced by them are within the coded picture area.

3.2.3 Loop filter

The prediction process may be modified by a two-dimensional spatial filter.

The filter characteristics are for further study. [Existing flexible hardware spec? What to do at block boundaries?]

The filter may be switched on or off on a block by block basis. The method of signalling this is for further study. [Explicit in TYPE3 or implicit from motion vector? Chrominance blocks?]

3.2.4 Transformer

Transmitted blocks are coded with a 2-dimensional transform of size 8 by 8. The input to the forward transform and output from the reverse transform have 9 bits. The output from the forward and input to the reverse have 12 bits.

The arithmetic procedures for computing the transforms are for further study.

3.2.5 Quantisation

The number of quantisers, their characteristics and their assignment are for further study.

3.2.6 Classification

Luminance block coefficients may be transmitted in any one of ... sequence orders as indicated by CLASS.

Chrominance block coefficients are transmitted in one sequence order.

The sequence orders are for further study.

3.2.7 Clipping

To prevent quantisation distortion of transform coefficient amplitudes causing arithmetic overflow in the encoder and decoder loops, clipping functions are inserted. In addition to those in the inverse transform, a clipping function is applied at both encoder and decoder to the reconstructed picture which is formed by summing the prediction and the prediction error as modified by the coding process. This clipper operates on resulting pel values less than 0 or greater than 255, changing them to 0 and 255 respectively.

3.3 Data rate control

Sections where parameters which may be varied to control the rate of generation of coded video data include processing prior to the source coder, the quantiser, block significance criterion and temporal subsampling. The proportions of such measures in the overall control strategy is not subject to recommendation.

When invoked, temporal subsampling is performed by discarding complete pictures. Interpolated pictures are not placed in the picture memory.

3.4 Forced updating

This function is achieved by forcing the use of the intra mode of the coding algorithm. The update interval and pattern are not subject to recommendation.

4. Video Multiplex Coder

4.1 Data Structure

Note: Unless specified otherwise the most significant bit is transmitted first.

Note: Unless specified otherwise all unused or spare bits are set to zero.

149

4.2 Video Multiplex arrangement

4.2.1 Picture Header

The structure of the Picture Header is shown in figure x.[based on fig.5 of #249] All Picture Headers are transmitted. When a picture is dropped by the encoder its Picture Header is followed immediately by the next Picture Header.

Picture Start Code (PSC) A unique word of 21 bits which cannot be emulated by error free data. Its value is ..... [0000 0000 0000 0001 1001 1 is strongly preferred by European countries because it is in sequence with the Group Numbers in the GBSC.]

Buffer state (BS) 6 bits representing the encoder buffer fullness, sampled at the top of picture. The unit of measure is for further study. [This definition differs from #249 because final buffer size is not yet determined and may vary with bit-rate.]

Temporal reference (TR) A three bit number derived using modulo-8 counting of Picture Start Codes.

Type information (TYPE1) Information about the complete picture

Bit 1 Split screen indicator. '0' off, '1' on.
Bit 2 Document camera. '0' off, '1' on.
Bits 3 to ... Spare. For further study.

Extra Insertion information (PEI)

For further study. [Some changes may be proposed here. For example is PGMV wanted? Should PARITY and PSPARE bit positions always be present and PEI bits signal that they contain useful information? Permanently allocated slots allow these bits to be inserted or modified in equipment after the encoder.]

Picture Global Motion Vector (PGMV)

For further study. [Is facility of genuine use? Are restrictions needed on the 16 bits to avoid emulation of start codes?]

Parity information (PARITY)

For optional use. 8 parity bits each representing odd parity of the aggregate of the corresponding bit planes of the locally decoded PCM values of Y, CR and CB in the previous picture period. In the Picture Header of a dropped picture, the parity bits are a repeat of those in the previous header.

Spare Information

For further study.

[Number of such bits and whether the positions are always inserted or not]

150

4.2.2 Group of blocks header

[Much of the Group of Blocks detail would be changed if a MACRO-BLOCK configuration is adopted]

A group of blocks consists of two lines of 44 luminance blocks each, one line of 22 CR blocks and one line of 22 CB blocks.

The structure of the Group of Blocks Header is shown in figure x. [based on fig.7 of #249] All GOB Headers are transmitted except those in dropped pictures.

Group of Blocks Start Code (GBSC) A word of 16 bits, 0000 0000 0000 0001.

Group Number (GN) A 5 bit number indicating the vertical position of the group of blocks. GN ranges from 00001 at the top of the picture to 10010 at the bottom of the picture.

Note: GBSC plus the following GN is not emulated by error-free video data.

Type information (TYPE2)

TYPE2 is ... bits which give information about all the transmitted blocks in a group of blocks.

Bit 1. When set to '1' indicates all 132 blocks are transmitted in intra mode.
Bits 2 to ?. Spare for further study.

[Functions of bits 2 and 3 in flexible hardware spec are omitted as they do not appear to serve any purpose which is not also covered by TYPE3. Is this correct?]

[TYPE2 may be a suitable place to carry the Freeze Picture Release for multipoint working.]

Quantiser information (QUANT1)

For further study.

[Number of bits depends on number of quantisers. First bit of flexible hardware spec would be omitted if QUANT2 never used. Should QUANT1 have two parts - one for INTER and one for INTRA?]

Extra insertion information (GEI)

For further study. [similar situation as with PEI.]

Group of Blocks Global Motion Vector (GGMV)

For further study. [similar comments as with GGMV)

Spare Information (GSPARE)

For further study. [Number of bits and uses. Consolidation with the spare bits in TYPE2?]

4.2.3 Block data alignment

The structure of the data for each transmitted block is shown in figure x.[similar to fig.8 of #249. Exact order may be optimised for error recovery and resilience once VLCs are known.] Elements are omitted when not required.

Block Address (BA) indicates the position of a block within a group of blocks. It is a Variable Length Code-word giving the run-length of non-transmitted blocks between this block and the previous transmitted block. In the first transmitted block in a GOB, the run-length is numerically equal to the absolute block address. VLC codewords are for further study.

For the purpose of run-length calculation, the block is regarded as continuous beginning with the upper line of luminance blocks through the lower ones, through CR and ending with CB blocks.

The ranges of absolute block addresses for the upper and lower lines of luminance blocks are 0 to 43 and 44 to 87 respectively. The ranges of absolute block addresses for the CR and CB lines of chrominance blocks are 88 to 109 and 110 to 131 respectively.

When all blocks in a GOB are coded in intra mode, BA is not transmitted for those blocks.

Block Type Information (TYPE3)

A variable length codeword indicating the type of block and which data elements are present. VLC codewords are for further study.

[Drafting of this part is left until decisions are made concerning such items as loop filter signalling, coding of motion vectors, retention of QUANT2.]

Quantiser (QUANT2) A fixed length code of ... bits signifying the table(s) employed to quantise the transform coefficients. QUANT2 is not present when the first bit of QUANT1 in the Group Header is set to '1'. [Should the facility to change the quantiser each block be retained?]

Classification Index (CLASS) A Variable length codeword present in luminance blocks specifying the transmission sequence of transform coefficients. The codewords are for further study.

Motion Vector Data (MVD)

Calculation of the vector data is for further study. [First or second modes. Macro blocks?]

When the vector data is zero, this is signalled by TYPE3 and MVD is not present.

When the vector data is non-zero, MVD is present consisting of a variable length codeword for the horizontal component followed by a variable length codeword for the vertical component.

Variable length coding of the vector components is for further study.

Transform coefficients (TCOEFF)

The quantised transform coefficients are sequentially transmitted according to the sequence defined by the class. The DC component is always first. Coefficients after the last non-zero one are not transmitted.

The coding method and tables are for further study.

End of Block Marker (EOB)

Immediately follows TCOEFF if present. Codeword for EOB is for further study.

4.3 Multipoint considerations

4.3.1 Freeze picture request

Causes the decoder to freeze its received picture until a picture freeze release signal is received. The transmission method for this control signal is for further study.

4.3.2 Fast update request

Causes the encoder to empty its transmission buffer and encode its next picture in intra mode with coding parameters such as to avoid buffer overflow. The transmission method for this control signal is for further study.

4.3.3 Data continuity

The protocol adopted for ensuring continuity of data channels in a switched multipoint connection is handled by the message channel. For further study.

5. Video Data Buffering

The size of the transmission buffer at the encoder and its relationship to the transmission rate are for further study.

6. Transmission coder

[Note: Consideration will be given to transferring some parts of this section to H.222 and AV.312 in subsequent versions]

6.1 Bit rate

The net bit-rate including audio and optional data channels is an integer multiple of 384 kbit/s up to and including 1920 kbit/s.

The encoder output clock rate shall be switchable between either a free running internal source or a source synchronised to the bit-stream received from the network. The mechanism for this switching is for further study.

153

When in free-running mode the tolerance on output clock rate will be ±50 ppm.

When in synchronised mode the synchronism should be maintained when the frequency of the received bit-stream is within ±50 ppm of nominal.

6.2 Video clock justification

Video clock justification is not provided.

6.3 Frame structure

6.3.1 Frame structure for 384-2048kbit/s channel

The frame structure is defined in CCITT Rec. H.222

6.3.2 Bit assignment in application channel
For further study.

6.3.3 Timeslot positioning according to I.431.

6.4 Audio coding

G.722 56/48 kbit/s audio, 0/8 kbit/s data and 8 kbit/s service channel in the first timeslot.

The delay of the encoded audio relative to the encoded video at the channel output shall be within the range -35 ms to +70 ms when the encoder buffer is empty.

6.5 Data transmission

One or more timeslots may be allocated as data channels of 64 kbit/s each. The first channel uses the fourth timeslot.

Positioning of the other channels, and possible restrictions on availability at lower overall bit rates are for further study. The BAS codes used to signal that these data channels are in use are specified in Rec. H.221.

[Note: 64kbit/s interfaces are usually X.21 which has a common clock for both directions of transmission. This is in conflict with the free-running mode of section 6.1 above.]

6.6 Error handling

For further study.

6.7 Encryption

For further study.

6.8 Bit Sequence Independence Restrictions

[Method to overcome restriction on bit sequence independence in North American network.] For further study.

6.9 Network interface

Access is at the primary rate with vacated timeslots as per I.431.

For 1544 kbit/s interfaces the default H0 channel is timeslots 1 to 6. Interface code is AMI or B8ZS.

For 2048 kbit/s interfaces the default H0 channel is timeslots 1-2-3-17-18-19. Interface code is HDB3.

END