5i’

2 Codecs not requiring separate television standards conversion when used on interregional connections

A codec for 525-line, 60 fields/s and 1544 kbit/s transmission for intra-regional use and capable of interworking with the codec of § 1

2.1 *Introduction*

Section 2 indicates the changes and additions which must be made to the text of § 1 in order to define the version of the codec for use with 525-line, 60 fields/s television standards and transmission at 1544 kbit/s. The two versions are capable of interworking via a re-multiplexing unit which can convert the Recommendation G.704, § 2.1 compatible frame structure on one side to the Recommendation G.704 , § 2.3 compatible frame structure (with 6 time slots empty) on the other side.

The two versions of the codec are identical in most respects, the important differences (apart from the obvious ones arising from different input and output signals) being confined to the digital pre- and post-filters and the signals for the control of the buffers. Moreover, the detailed algorithms of the pre- and post-filters do not need to be specified to permit interworking. Only an outline of their mode of operation together with the few necessary specifications are therefore provided.

2.2 *Brief specification*

2.2.1 *Video input/output*

The video input and output are standard 525-line, 60 fields/s colour or monochrome television signals. The colour signals are in component form. Colour and monochrome operation are fully compatible.

2.2.2 *Digital output/input*

The digital output and input are at 1544 kbit/s, compatible with the frame structure of Recommendation G.704.

2.2.3 *Sampling frequency*

The video sampling frequency and 1544 kbit/s network clock are asynchronous.

2.2.4 *Coding techniques*

Conditional replenishment coding supplemented by adaptive digital filtering, differential PCM and variable-length coding are used to achieve low bit-rate transmission.

2.2.5 *Audio channel*

An audio channel using 64 kbit/s is included. At present, coding is A-law according to Recommendation G.711, but provision is made for future use of more efficient coding.

2.2.6 *Mode of operation*

The normal mode of operation is full duplex.

2.2.7 *Codec-to-network signalling*

An optional channel for codec-to-network signalling is included.

2.2.8 *Data channels*

Optional 2×64 kbit/s and 1×32 kbit/s data channels are available. These are used for video if not required for data.

2.2.9 *Forward error correction*

Optional forward error correction is available. This is required only if the long-term error rate of the channel is worse than 1 in 10^6.

2.2.10 *Additional facilities*

Provision is made in the digital frame structure for the future introduction of encryption, a graphic mode and multipoint facilities.

2.2.11 When the coder buffer is empty and the decoder buffer full, the coder delay is 31 ± 5 ms and the decoder delay is 176 ± 31 ms

2.3 *Video interface*

The normal video input is a 525-line, 60 fields/s signal in accordance with CCIR Report 624. When colour is being transmitted, the input (and output) video signals are in component form. The luminance and colour-difference components, E'_Y, $(E'_R — E'_Y)$ and $(E'_B — E'_Y)$ are as defined in CCIR Report 624. The video interface is as recommended in CCIR Recommendation 567.

2.4 *Source coder*

2.4.1 *Luminance component or monochrome*

2.4.1.1 *Analogue-to-digital conversion*

The signal is sampled to produce 256 picture samples per active line (320 samples per complete line). The sampling pattern is orthogonal and line, field and picture repetitive. For the 525-line input, the sampling frequency is 5.0 MHz, locked to the video waveform.

Uniformly quantized PCM with 8 bits/sample is used.

Black level corresponds to level 16 (00010000).

White level corresponds to level 239 (11101111).

PCM code words outside this range are forbidden (the codes being used for other purposes). For the purposes of prediction and interpolation, the final picture element in each active line (i.e. picture element 255) is set to level 128 in both encoder and decoder.

In all arithmetic operations, 8-bit arithmetic is used and the bits below the binary point are truncated at each stage of division.

2.4.1.2 *Pre- and post-filtering*

These are typical figures. The delays depend upon the detailed implementation used.

2.4.1.2.1 *Spatial filtering*

A digital filter reduces the 242½ active lines-per-field of the 525-line signal to 143 lines-per-field, the same number as in the 625-line version of the codec. In the decoder, the digital post-filter uses interpolation to restore the signal to 525-lines per picture.

2.4.1.2.2 *Temporal filtering*

A recursive temporal pre-filter with non-linear transfer characteristics is used in the coder to reduce noise in the signal and increase coding efficiency. The frame store used in this filter can also be used as the storage element of a frame interpolator with variable coefficients which is used to reduce the transmitted frame rate to a value less than that of the input video signal. In 525-line to 525-line transmission, the transmitted frame frequency is locked to the video clock and is approximately 29.67 Hz (29.97 Hz times 3057/3088) instead of the nominal video rate of 29.97 Hz. In 525-line to 625-line transmission, the transmitted frame frequency is nominally 25 Hz and is locked to the channel clock.

Because the (television) frames are leaving the coder more slowly than they are entering, the coding process is suspended for one frame every *N* th input frame. *N* is approximately 100 for 525-line to 525-line operation and approximately 6 for 525-line to 625-line operation.

In the decoder, the digital post-filter incorporates a frame store in some versions of the 625-line codec where it is used in the line interpolation process. In the 525-line version, in addition to its use for line interpolation, it is used as a temporal interpolator with variable coefficients to provide an extra output frame during those periods when the decoding is temporarily suspended.

2.5 *Video multiplex coding*

2.5.1 *Buffer store*

The size of the buffer store is defined at the transmitting end only and is 160 kbits. Of this, 96 kbits is used for smoothing the video data in the face-to-face mode and the remainder is used to accomodate the action of the frame interpolator (see § 2.5.1.1 below) and the requirements of the graphics mode.

At the receiving end, the buffer must be at least this length but in some implementations of the decoder, it may be longer.

2.5.1.1 *Buffer control*

The amount to which the transmitting buffer is filled is used to control various coding algorithms (subsampling, etc.) and is signalled to the decoder to enable it correctly to interpret the received signals. In the 525-line codec, the transmission rate is less than the video input rate and hence the buffer tends to fill more rapidly than would be determined by the movement in the picture, only to empty again when the interpolator suspends the coding process.

To avoid incorrect changes in coding algorithms, the buffer-state signal is modified to take account of the progressively changing coefficients of the interpolator in the pre-filter. The buffer then operates as though the data is coming from a video source whose frame rate is uniform and the same as the transmitted frame rate.

2.6 *Transmission coding*

The transmission coder assembles the video, audio, signalling and optional data channels into a 1544 kbit/s frame structure which is compatible with Recommendation G.704.

2.6.1 *Serial data*

See § 1.6.1.

2.6.2 *Audio*

See § 1.6.2.

2.6.3 *Transmission framing*

The frame structure, compatible with Recommendation G.704 and also compatible with that of the 625-line version in § 1, is given in § 2 of Recommendation H.130.

2.6.3.1 *General*

See § 1.6.3.1.

2.6.3.2 *Use of certain bits in each octet in the odd frames of time slot 2*

The use of certain of the bits in time slot 2 (odd) differs slightly from that given for the codec in § 1. The differences are as follows:

Bit 1 — For clock justification

This bit is disregarded in 525-line decoders.

To permit interworking with the 626-line codecs of § 1, the 525-line coders must transmit a fixed bit-pattern which is used to control the frequency of the video clock in 625-line decoders. The exact form of the repetitive pattern need not be specified but it must contain seven ‘‘ones’’ and four ‘‘zeros’’ in 11 bits, e.g.:

1 0 1 1 0 1 0 1 1 0 1

Bit 2 — To signal buffer state

The degree to which the encoder buffer is filled, after correction for the interpolator (see § 2.5.1.1), is measured in increments of 1 K (1 K = 1024 bits), and signalled using an 8-bit binary code. When working to a 525-line decoder, the buffer state is sampled every 3057 channel-clock periods. When working to a 625-line decoder, the buffer state is sampled 10 times during every 525-line field period. When the buffer input is suspended for a frame period, the buffer sampling is stopped. The sampled values of the buffer state are stored prior to transmission. The store may hold between zero and 23 values which have been modified to take account of the interpolator coefficients at the times of sampling. The modified sample values are read out [as bit 2 of TS2 (odd)] at a uniform rate; the most significant bit (MSB) in frame 1 of the multiframe, the second MSB in frame 2, etc.

Bit 3.7 — Fast update request

On receipt of this bit set to 1, the transmitter buffer is forced to decrease its full and stabilise to a modified state of less than 6 K by preventing coded picture elements from entering the buffer. Bit A is set to 1 in the next FST. The two following fields are treated as complete moving areas and the encoder uses an arrangement for control of the sub-sampling modes to make the buffer overflow condition unlikely.

3 A codec for 525-lines, 60 fields/s and 1544 kbit/s transmission for intra-regional use

3.1 *Introduction*

A 1.5 Mbit/s interframe codec described under § 3, is capable of transmitting and receiving a single NTSC video signal and audio signal using an adaptive predictive coding technique with motion-compensated prediction , background prediction and intraframe prediction

The aim of this codec is to effectively transmit video telephone and video conferencing signals which have relatively small movements. The video interface of the codec is a 525-line, 60 fields/s standard analogue television signal corresponding to the ‘‘Class *a* ’’ standard of Recommendation H.100.

3.2 *Outline of codec*

The essential parts of the codec block diagram are shown in Figure 7/H.120. The coder consists of three basic functional blocks, that is, pre-processing, video source coding and transmission coding.

In the pre-processor, the input analogue NTSC video signal is digitized and colour decoded into one luminance component and two chrominance components. These three components are time division multiplexed into a digital video form, whose noise and unnecessary signal components are removed by the pre-filter.

In the video source coder, the digital video signal is fed to the predictive coder where interframe and intraframe predictive coding techniques are fully utilized for minimizing prediction errors to be transmitted. The prediction error signal is next entropy-coded using its statistical properties to reduce redundancies. Since the coded error information is generated in irregularly spaced bursts, a buffer is used. If the buffer becomes full, the number of prediction error quantizing levels and/or picture elements to be coded is reduced to prevent any overflow.

In the transmission coder, coded video and audio signals are first encrypted on an optional basis. The coded video signal is then forward error correction coded and scrambled. The three signals, coded video, coded audio and optional data signals are multiplexed into a 1544 kbit/s digital format with a frame structure as defined in Recommendation H.130.

The decoder carries out a reverse operation.

Figure 7/H.120, p.

3.3 *Brief specification*

3.3.1 *Video input/output*

NTSC signals are used for the video input/output signal, with monochrome signals being additionally applicable.

3.3.2 *Digital output/input*

The interface conditions for the digital output/input signal satisfy Recommendation G.703 specifications. The signal transmission rate is 1544 kbit/s.

3.3.3 *Sampling frequency*

The video sampling frequency is four times the colour sub-carrier frequency ($f_{S\backslash dC}$) and asynchronous with the 1544 kHz network clock.

3.3.4 *Time division multiplexed (TDM) digital video format*

An NTSC signal is separated into a luminance component (Y) and two chrominance components (C_1 and C_2). A time division multiplexed signal composed of Y and time-compressed C_1 and C_2 is employed in the source coding as the standard digital video format.

3.3.5 *Coding algorithm*

Adaptive predictive coding supplemented by variable word-length coding is used to achieve low bit rate transmission. The following three predictions are carried out adaptively on a pel -by-pel basis:

a) motion-compensated interframe prediction for a still or slowly moving area,

b) background prediction for an uncovered background area, and

c) intraframe prediction for a rapidly moving area.

Prediction errors for video signals and motion vectors are both entropy-coded using the following two techniques:

i) variable word-length coding for non-zero errors, and

ii) run-length coding for zero errors.

3.3.6 *Audio channel*

An audio channel using 64 kbit/s is included. The audio coding algorithm complies with Recommendation G.722.

3.3.7 *Data channel*

An optional 64 kbit/s data channel is available, which is used for video if not required for data.

3.3.8 *Mode of operation*

The normal mode of operation is full duplex, with other modes, e.g. the one-way broadcasting operation mode, also taken into account.

3.3.9 *Transmission error protection*

A BCH error correcting code is used along with a demand refreshing method to prevent uncorrected errors from degrading the picture quality.

3.3.10 *Additional facilities*

Provision is made in the digital frame structure for the future introduction of such facilities as encryption, graphics transmission and multipoint communication.

3.3.11 *Processing delay*

The coder plus decoder delay is about 165 ms without that of a pre-filter and a post-filter.

3.4 *Video interface*

The video input/output signal of the codec is an analogue NTSC signal (System M) in accordance with CCIR Report 624.

3.5 *Pre- and post-processing*

3.5.1 *Analogue-to-digital and digital-to-analogue conversion*

An NTSC signal band-limited to 4.5 MHz is sampled at a rate of 14.3 MHz, four times the colour sub-carrier frequency ($f_{S\backslash dC}$), and converted to an 8-bit linear PCM signal. The sampling clock is locked to the horizontal synchronization of the NTSC signal. Since the sampling frequency is asynchronous with the network clock, the justification information is coded and transmitted from the coder to the decoder.

The digital video data is expressed in two's complement form. The input level to the A/D converter is defined as follows:

— sinc tip level (—40 IRE) corresponds to —124 (10000100);

— white level (100 IRE) corresponds to 72 (01001000).

(IRE: Institute of Radio Engineers)

As a national option, a pad can be inserted before the A/D converter if a level fluctuation should be taken into account at analogue transmission lines connecting terminal equipment and codec.

At the decoder, the NTSC signal is reproduced by converting the 8-bit PCM signal to an analogue signal.

3.5.2 *Colour decoding and encoding*

The digitized NTSC signal is separated into the luminance component (Y) and the carrier band chrominance component (C) by digital filtering. The two baseband chrominance signals (C_1 and C_2) are obtained by digitally demodulating the separated carrier band chrominance component. The effective sampling frequency after colour decoding is converted to 7.2 MHz (2 $f_{S\backslash dC}$) and 1.2 MHz (1/3 $f_{S\backslash dC}$) for the luminance signal and chrominance signals respectively.

The replica of the NTSC signal is obtained by digitally modulating the C_1 and C_2 signals and adding to the Y signal at the decoder.

Filter characteristics for colour decoding and encoding are left to each hardware implementation since they do not affect interworking between different design codecs. Examples of recommended characteristics are described in Annex E.

3.5.3 *TDM signal*

A time division multiplexing (TDM) signal is constructed from the separated component signals.

First, the C_1 and C_2 signals are time-compressed to 1/6. Next, each of the time compressed C_1 and C_2 signals, with their horizontal blanking parts removed, is inserted into the Y signal horizontal blanking interval on alternate lines. C_1 is inserted on the first line of the first field and on every other line following throughout the frame, while C_2 is inserted on the second line of the first field and on every other line following throughtout the frame.

Active samples for the Y signal are 384 samples/line and 64 samples/line for the C_1 and C_2 signals. The TDM signal is constructed with these active samples and 7 colour burst samples (B), which are inserted into the top of the TDM signal.

As shown in Figure 8/H.120, the C_1 and C_2 signal sampling points coincide with that of the Y signal on every sixth sample. The C_1 and C_2 signals of only the odd lines are transmitted to the decoder.

At the decoder, each component signal is again demultiplexed from the TDM signal, and time-expansion processing of 6 times is carried out for the C_1 and C_2 signals.

Note — When a pad is inserted before the A/D converter as described in § 3.5.1, pre-emphasis (de-emphasis) with a compensating gain for the $C_{1,\ 2}$ and colour burst signals is recommended at the source coder input (decoder output) to obtain better picture reproduction in coloured parts.

3.5.4 *Pre- and post-filtering*

In addition to conventional anti-aliasing filtering prior to analogue-to-digital conversion, the following two filtering processes should be used as pre-filtering for source coding:

a) temporal filtering to reduce random noise included in the input video signal;

b) spatial filtering to reduce aliasing distortion in subsampling.

At the decoder, the following three filtering processes should be used as post-filtering in addition to conventional low pass filtering after digital-to-analogue conversion:

i) spatial filtering to interpolate the omitted picture elements in subsampling;

ii) spatio-temporal filtering to interpolate the omitted fields in field repetition;

iii) temporal filtering to reduce noise generated in the course of source coding.

Although these filtering processes are important for improving reproduced picture quality, their characteristics are independent of interworking between different design codecs. Hence, pre- and post-filtering is left to each hardware implementation.

Figure 8/H.120, p.

3.6 *Source coding*

3.6.1 *Configuration of source coder and decoder*

The video source coder and decoder configuration of this codec is outlined in Figure 9/H.120.

The predictive encoder converts the input video signal *x* into the prediction error signal *e* , using the motion vector *v* . This conversion is controlled by the coding mode *m* .

The variable word-length (VWL) coder codes *e* and *v* into the compressed data *C* using the variable length coding method. The transmission buffer memory (BM) smoothes out the irregularly spaced data *C* . The coding mode *m* is also coded.

The frame memory parity information *p* is used to check the identity of coder and decoder frame memory contents. If any parity error is detected, frame memories of both coder and decoder are reset by the demand refresh information (DR) and the demand refresh confirmation information (DDR).

At the decoder, the variable word-length (VWL) decoder decodes *e* , *v* , *m* and *p* , and the predictive decoder reproduces the video signal *x* ‘.

Figure 9/H.120, p.

3.6.2 *Predictive coding*

3.6.2.1 *Coding modes*

Five coding modes as summarized in Table 3/H.120 are provided. All of the samples are coded and transmitted in normal mode, while half of the samples are omitted in subsampling mode. In field repetition mode, one or more consecutive fields are omitted (called multi-field repetition, see Note 1). If field repetition mode and subsampling mode are used in combination, only a quarter or less of the original picture elements are coded and transmitted.

Subsampling is carried out in a quincunx way, namely by transmitting only odd-numbered pels on odd-numbered lines and even-numbered pels on even-numbered lines in each block-line (see Note 2).

In field repetition mode, either the odd or even fields are omitted. For the omitted fields, both the prediction error *e* and the motion vector *v* are set to 0.

Note 1 — If odd fields and even fields are mixed after field omission, a severe picture degradation takes place. Hence, 1 out of 2, 3 out of 4 or 5 out of 6 field omission is recommended.

Note 2 — Each block-line consists of 8 lines as defined in § 3.6.2.5.

H.T. [T7.120]
TABLE 3/H.120
Coding modes

Coding modes	Abbreviation	Operation	
1	Normal	NRM	Full sampling
2	Field repetition	FRP	One or more fields omission
3	Subsampling	SBS	2: 1 per omission
4	Stop	STP	Suspension of coding
5	Refresh	RFS	Renewal of frame memory

Tableau 3/H.120 [T7.120], p.

3.6.2.2 *Adaptive prediction*

Prediction functions are adaptively selected on a pel-by-pel basis as shown in Figure 10/H.120. The selection is carried out so as to minimize probable prediction errors. This is accomplished using the two prediction status signals, which are determined by prediction reference signals, for the preceding pels located on the previous and the present lines.

When subsampling and/or field repetition are operated, omitted pels are interpolated in the prediction loop.

The notations defined for the i-numbered pel are as follows:

X_i | local decoder output,

Y_i | interpolator output,

M_i | motion compensated interframe prediction value,

B_i | background prediction value,

I_i | intraframe prediction value,

* | logical product, and

+ | logical sum.

Figure 10/H.120, p.

3.6.2.2.1 *Motion-compensated interframe prediction/background prediction*

Prediction status signal S_{1i} | for pel i
| is determined as

where prediction reference signal $R_1(i)$ is

Based on S_{1i}, prediction signal X_{1i} | is given as

If pel i | is either omitted due to subsampling and/or field repetition or forced intraframe coded or in burst B , its corresponding $R_i(i)$ is set to 0 regardless of equation (3-2).

3.6.2.2.2 *Interframe prediction/intraframe prediction*

Prediction status signal S_{2i} | for pel i
| is determined as

where prediction reference signal $R_2(i)$ is

Based on S_{2i}, prediction signal X_{2i} | is given as

If pel (i — 1) is omitted due to subsampling, $R_2(i — 2)$ is used instead of $R_2(i — 1)$. On the other hand, if pel (i — 455) is omitted, $R_2(i — 454) * R_2(i — 456)$ is used instead of $R_2(i — 455)$. If pel i is forced intraframe-coded, its corresponding $R_2(i)$ is set to 1 regardless of equation (3-5).

If pel i | is omitted due to field repetition, its corresponding $R_2(i)$ is set to 0 regardless of equation (3-5). When pel i is not forced-intraframe coded, $R_2(i)$ in burst B is set to 0.

3.6.2.3 *Background generation*

The background prediction value is generated scene adaptively as

where
[Formula Deleted]