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

Document #141
Sep. 12, 1986

SOURCE: CHAIRMAN OF THE SPECIALISTS GROUP ON CODING FOR VISUAL TELEPHONY

TITLE : SPECIFICATIONS FOR REFERENCE MODEL VERSION 2 (RM2)

In order to investigate the coding algorithm, Reference Model was agreed in Tokyo (see Document #104). Several improvements were added to this model in Montreal to define Reference Model Version 2 (see Annex 1 to Document #140R). Since the following two modifications have been found necessary through preliminary experiments, they are also included in RM2.

- Intra/inter mode decision should prevent intra mode selection when prediction error power is small
- Quantizer step size should be controlled more finely in lower step size region

This document describes whole of the RM2 specifications as a revision of Document #104, with the following indications in the left margin.

- Single solid line for items agreed in Montreal (Annex 1 to Document #140R)
- Double solid line for the above two items
- Zigzag line for clarification

As for the data presentation method, Chairman received suggestions from Mr. Brusewitz to clarify Section 1/Annex 1 to Document #140R. These are also included in this document.

---

1) The reference simulation model has a configuration as shown in Figure 1.

2) The same sequences at the Common Intermediate Format must be used: 'Miss America', 'Checked Jacket', 'Split-Trevor' and 'Graphics'.

2.1) When presenting processed pictures of Split-Trevor, it is suggested to have ;

a) the complete sequence
b) only the Split part
c) only the Trevor part

3) Number of coded pels per line is reduced to 352 by discarding the first 8 pixels.

4) Classical DCT (a) of Temporary Document No. 3 (Tokyo).

5) When we count the information rate, only number of bits necessary for transmitting block attributes and block data are calculated. Those bits necessary for picture attributes and group of block attributes are neglected, because they are only several tens bits and small enough compared with the average number of bits per picture (20 or 30 kbits corresponding to 2:1 or 3:1 temporal subsampling) or per group of block (1111 or 1666 bits).

6) Block types are defined as shown in Figure 2. A block type is determined in the following order.

1. MC/No MC : according to Item 21)
2. Intra/Inter : according to Item 14)
3. Coded/Non coded : according to Item 8)

6.1) The quantizer step size is transmitted as a GOB attribute.

7) Number of bits per block required to transmit each block type are shown in Table 1, where the block attribute 'fpvt' to identify the block type is represented with variable length codes defined in Paragraph 3 of p.1/Document #107.

8) Coded/Non coded Attribute value is determined after finding out whether the block includes one or more transform coefficients which are outside the range of - 1.5g to + 1.5g.

9) Buffer size and buffer state is written in kbits. 'k' represent 1000 in this case.

10) Block data is represented as follows (see Figure 3).

[Data] = q = [index(0,0)][index(0,1)][index(1,0)][index(2,0)] [index(1,1)]...[EOB]

Fig. 3

(0,0)	(0,1)	(0,2)	---	(0,7)
(1,0)	(1,1)	(1,2)	---	(1,7)
---	---	---	---	---
(7,0)	(7,1)	(7,2)	---	(7,7)

11) Diagonal scanning: see Annex 1.

12) Quantizer, VLC and EOB: see Annex 2. VLC code set in Table 4/ Document #122 is applied. EOB=010. Maximum values correspond to +/- 70th quantizer output. Clipping must be introduced for coefficients whose absolute value is greater than the 70th quantized output.

13) For the DC component of intra-frame coding, 9 bit linear quantization is used.

Note: Experience shows that fine quantization of the DC component plays an important role in getting good pictures.

14) Inter/intra modes and how to handle a scene cut: see Annex 3. It should be noted that a priori knowledge is used for scene cut.

15) The following temporal subsampling to CIF is used.

Miss America	2:1
Checked Jacket	2:1
Split-Trevor	3:1
Graphics	3:1

16) Buffer size: a size of 30 kbits is intended. However, for the reference model simulation purpose, the buffer fullness may exceed 30 kbits to avoid special anti-overflow measures (see the next item).

Note: It is expected that the present improvements of the model would not make the buffer fullness go over 30 kbits.

17) Buffer control: see Annex 4. It should be noted that coding is continued with g = 32 even when the buffer fullness reached 30 kbits.

18) Buffer fullness simulation: Since quantizer step size is controlled on a GOB-by-GOB basis, the buffer fullness B(n) is calculated at every GOB as follows.

$$B(n) = B(n-1) + I(n) - A/18$$

where B(n) : Buffer fullness at the end of the n-th GOB
B(n-1): Buffer fullness at the end of the (n-1)-th GOB
I(n) : Number of bits spent to transmit the n-th GOB (cf Table 1)
A : Average number of bits transmitted per picture
= 40000 for 2:1 temporal subsampling, scene cut
= 20000 for 2:1 temporal subsampling, other pictures
= 60000 for 3:1 temporal subsampling, scene cut
= 30000 for 3:1 temporal subsampling, other pictures

It should be noted that B(n) be calculated using 'real' type variable.

Note: If we would simulate the hardware codec more exactly, above value of A is not correct. For example, in case of '2:1 temporal subsampling, other pictures', it should be 10000 and we have to subtract 10000 at the end of each coded picture. However, for our reference model purpose, this strictness will not be necessary.

19) Motion estimation is achieved by block matching technique with a window search of $\pm$ 7.

- Integer displacement only.
- Coarse-fine 3 steps algorithm (see Annex 5) is used.

Stage 1: Search at 0 and $\pm$ 4 pels.
Stage 2: New center at minimum error position of stage 1. Search at 0 and $\pm$ 2 pels (referring to the new center).
Stage 3: New center at minimum error position of stage 2. Search at 0 and $\pm$ 1 pels (referring to the new center).

Search area is restricted to the inside of the 352 pels x 288 lines even when the block is located at the edge of the picture.

- Non-zero motion vectors are transmitted by using an 8 bit FLC.

20) No motion vector for chrominance.

21) Motion vectors are forced to zero according to Fig. 4, where the following two values are compared.

- $\sum^{block}\sum$ |frame difference| / 64
- $\sum^{block}\sum$ |displaced frame difference| /64

22) Evaluation function for motion estimation is a sum of absolute difference concerning to all of the 64 pels in a block.

22.1) The order of the search in each step is suggested in Annex 5. If the evaluation fuction becomes minimum at multiple displacements, the displacement with the least order number is selected.

23) No pre-processing, no post-processing is allowed.

24) No filtering in the loop is allowed.

25) Presentation of new algorithms or parameters: use the window method proposed in Annex 6. It is suggested, in order to minimize the amount of sequences to be processed, to use the most critical sequence for the comparison between the Reference and the improvement on the Reference. The most critical sequence may depend on the type of processing. The ultimate results should be shown, however, by presenting again all the sequences.

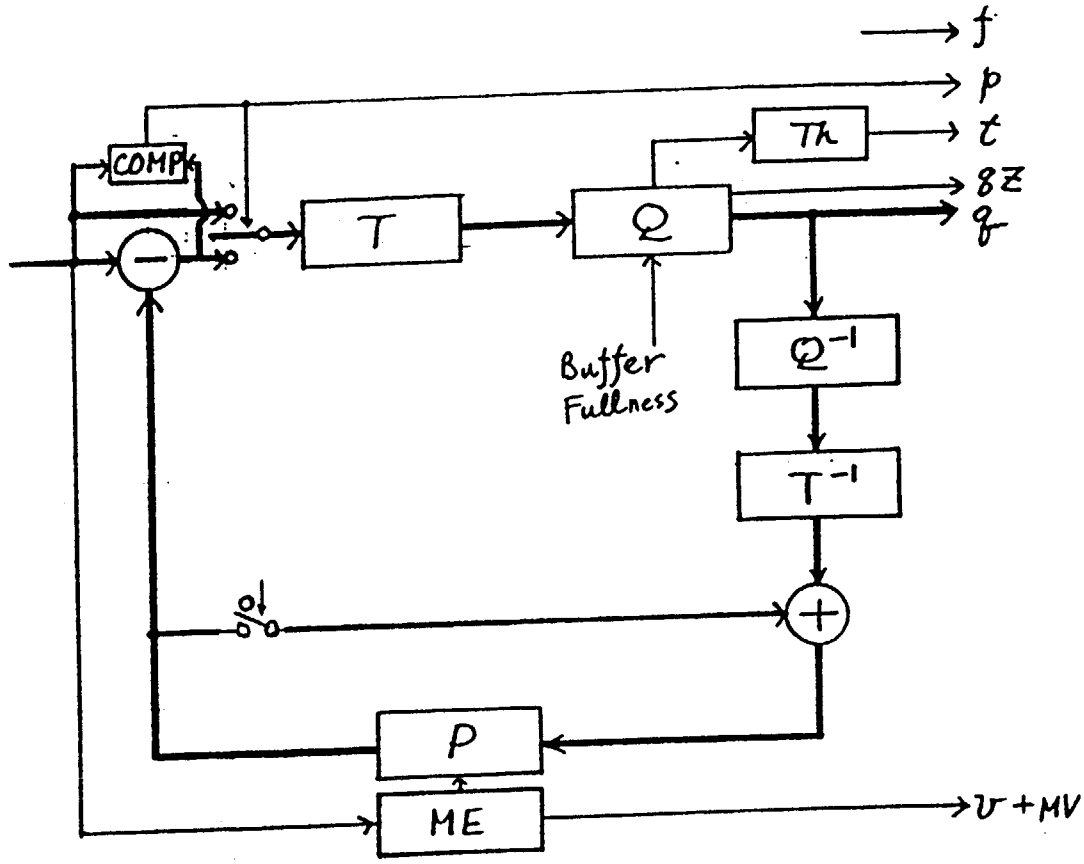
26) For presentation of the numerical data, the format in Annex 7 is suggested, which also includes the definition of each term. Data should be presented for both of the average of the whole sequence and the 15th encoded picture (Note).

Note:

Original picture	1	2	3	4	5	6	7	8	9	..
Coded picture	1		*		2		3		4	..

* omitted after the start

End



bs : Buffer State
f : Fixed / Non Fixed (Attribute ≠ 1 / Doc. #69)
v : No MC / MC (Attribute ≠ 2 / Doc. #69)
t : Non coded / Coded (Attribute ≠ 3 / Doc. #69)
p : Intra / Inter (Attribute ≠ 4 / Doc. #69)
gz : Quantizer indication
q : Quantizing index + EOB
MV: Motion Vector

Fig. 1 REFERENCE MODEL

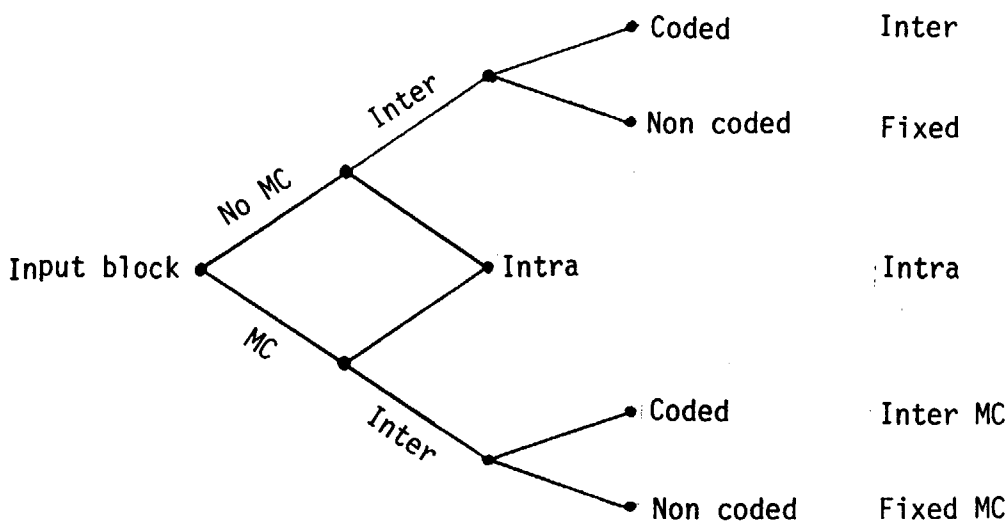


Figure 2 Block type

Table 1 Number of bits per block

Block Type	Block Attrib.			Quantiz. Index	Total	Note
	fpvt		MC Vector			
	Y	C				
Inter	3	2	0	[Data]	3(Y)/2(C) + [Data]	
Fixed	1	1	0	0	1	
Intra	4	2	0	[Data]	4(Y)/2(C) + [Data]	
Inter MC	2	-	8	[Data]	10 + [Data]	
Fixed MC	4	-	8	0	12	

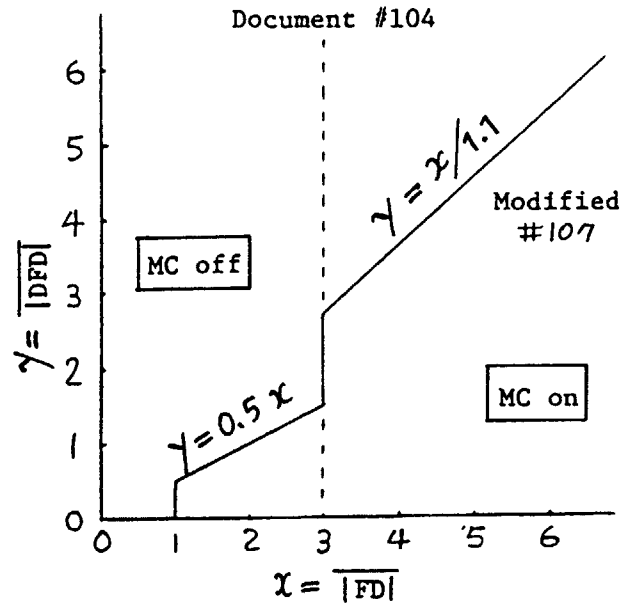
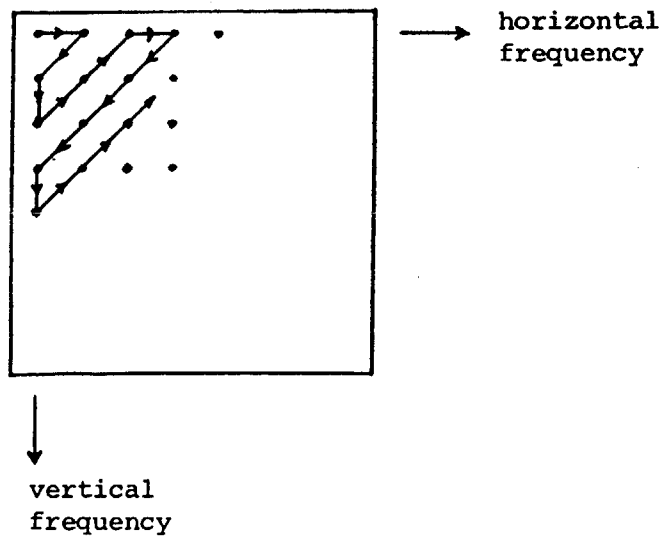


Fig. 4

Note: 'MC off' includes the solid line.

Annex 1: Diagonal Scanning

The transform coefficients are transmitted in the following way:



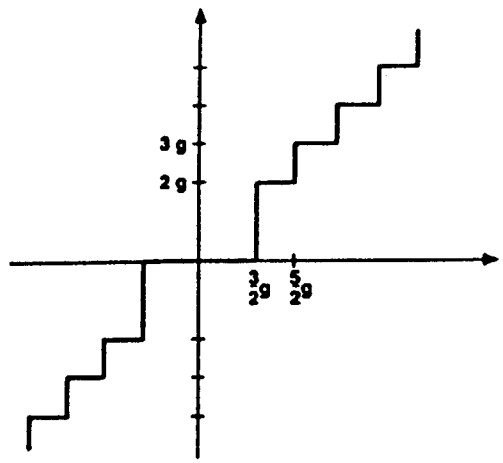
That means:

DC, first horizontal frequency, first vertical frequency, ...

Horizontal frequency corresponds to vertical lines in the picture.

Annex 2: Quantizer, VLC and EOB

Uniform quantizer is defined by the step size g. The quantizer has a threshold with value 1.5g.



Example: A transform coefficient F

$$1.5g \leq F < 2.5g$$

is quantized to the value $\hat{F} = 2g$.

The dynamic range is ± 2048. Step size is related to uniform transform, alternative (a)/Temporary Document No. 3 (Tokyo).

The same quantizer is used in both inter- and intra-frame coding and also in both luminance and chrominance coding. Quantized transform components are coded with the following VLC. An exception is in DC components of intra-frame mode, which are 9 bit linearly quantized and coded with a 9 bit FLC.

Quantizer Output	VLC Wordlength
0 g	1 bit
+/- 2 g	3 bits
+/- 3 g	5 bits
+/- 4 g	6 bits
+/- 5 g	7 bits
+/- 6 g	8 bits
+/- 7 g	9 bits
+/- 8 g	16 bits
+/- 9 g	16 bits
.	.
.	.
.	.
+/- 71 g	16 bits

Annex 3: Inter/intra Mode Switch and Scene Cut

1. Inter/intra-frame switch

a) Block based.
b) If we put;

$x(i)$: pel value of the input picture,
$\bar{x}$: average pel value over the block, and
$d(i)$: motion compensated interframe prediction error of the block,

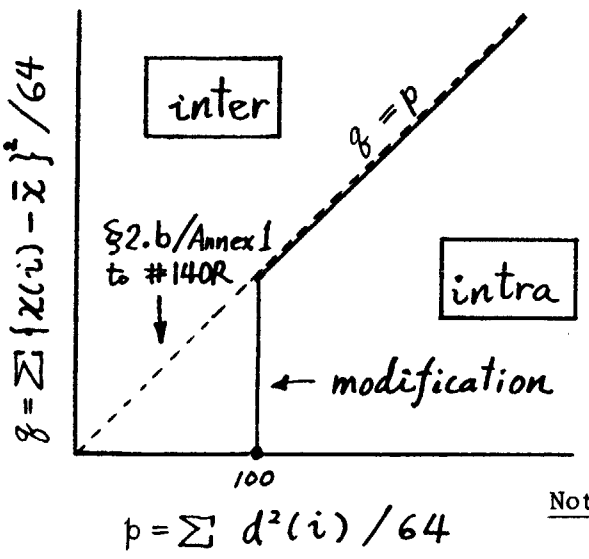
the intra/inter mode switch condition is a comparison of

$$\sum^{\text{block}} \{x(i)-\bar{x}\}^2/64 \quad \text{and} \quad \sum^{\text{block}} d^2(i)/64.$$

When both of them are small, intra mode selection is not appropriate even if the corresponding power is smaller, because

- intra mode block is always transmitted,
- at least 9 bits per block is necessary since DC component is coded in 9 bits (see Item 13),
- tile artifacts are liable to take place.

Hence inter mode should be selected in this case. A modification to the mode switch condition is shown in the following figure. This is provided for reference model purpose; optimum threshold or other effective methods should be studied further.



Note: 'Inter' includes the solid line.

2. Scene cut

A scene cut is given by a priori knowledge. In the case of scene cut, intraframe mode is used. To overcome the problem of a buffer overflow, the previous picture should be repeated, that means, for

the first picture after the scene cut, the amount of bits available is twice compared with the average amount of bits per picture.

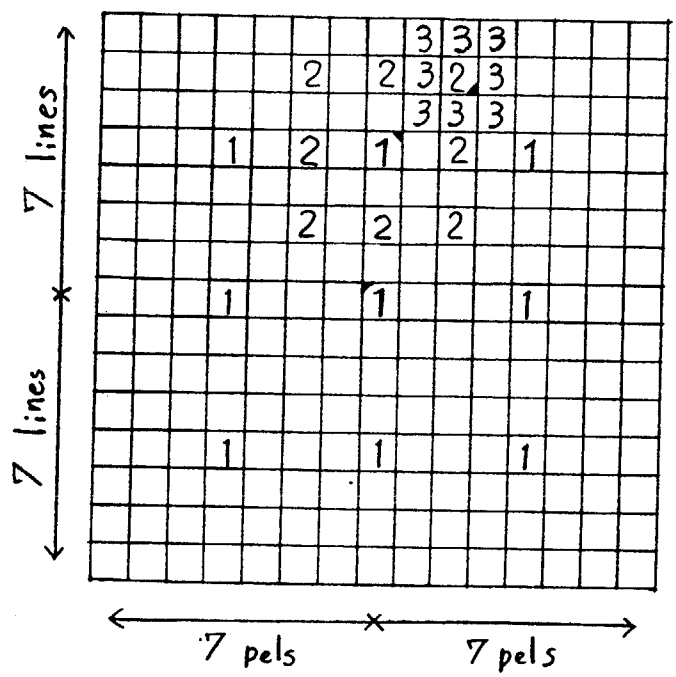
Annex 4: Buffer Control

Quantizer step size g is chosen for each GOB, and transmitted as a GOB attribute. Step size is chosen as a function of buffer content:

$$g = \begin{cases} 4, & \text{for } BF < 3000 \\ \text{INTEGER } (BF/1000) + 2, & \text{for } 3000 \leq BF < 30000 \\ 32, & \text{for } BF \geq 30000 \end{cases}$$

where g is step size and BF is buffer fullness (in bits).

Annex 5: Motion Estimation



1: first step
2: second step
3: third step

□: center of the first step

□: center of the second step

□: center of the third step

2 3 4
5 1 6
7 8 9

Order of the search in each step

Coarse-fine 3 steps algorithm

Annex 6: Presentation of Improvements

Take a 180x288 window from the "Reference Model"-result and take the same window from the improved result; put them next next to each other (so the results can be displayed together).
For the Split/Trevor-sequence and the Checked Jacket-sequence we propose to position the window at (40,0):

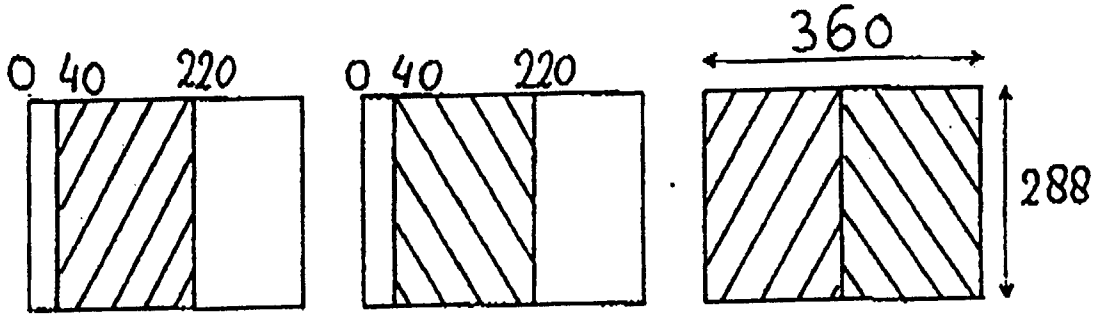


fig. 1 reference model result | improvement | refer. improv. DISPLAY

For the Miss America-sequence we propose to position the window at (90,0):

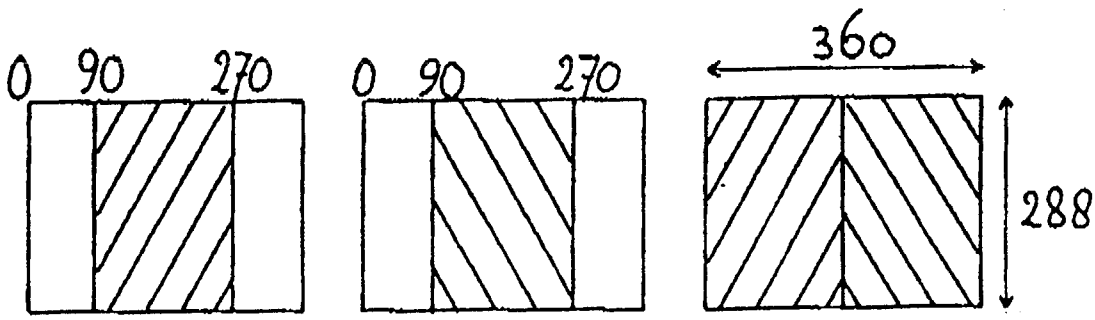


fig. 2 reference model result | improvement | refer. improv. DISPLAY

Note: In figure 1 and 2, the column 0 corresponds to the first column of the modified (352 pels x 288 lines) CIF.

Annex 7: Numerical Data Presentation Format

Sequence :

Item			Average of the whole sequence	The 15th encoded picture
1)R.M.S. for luminance				
2)SNR for luminance				
3)Mean value of the step size				
4)Mean value of the number of non-zero coefficients				
5)Mean value of the number of zeroes before the last non-zero coefficient				
6) Block type of Y	Intra			
	Fixed (Inter/No MC/No coded)			
	Inter (Inter/No MC/Coded)			
	Fixed MC (Inter/MC/No coded)			
	Inter MC (Inter/MC/Coded)			
7) Block type of C	Intra	C_R		
		C_B		
		C_R+C_B		
	Fixed (Inter/No coded)	C_R		
		C_B		
		C_R+C_B		
	Inter (Inter/Coded)	C_R		
		C_B		
		C_R+C_B		
8) Number of bits	Attributes	Y		
		C_R		
		C_B		
		Total		
	EOB words			
	Motion Vectors			
	Coefficients	Y		
		C_R		
		C_B		
		Total		
	Total			

Definitions

It should be noted that the first picture as well as the scene-cut picture (if any) are excluded when calculating

$$\text{'Average of the whole sequence'} = \frac{\sum^{seq} (\quad)}{\text{number of pictures}}$$

and that luminance only for 1), 2) while luminance and chrominance for 3), 4), 5).

1) RMS coding error

$$RMS_{PIC} = \text{sqrt}(MS_{PIC})$$
$$RMS_{SEQ} = \text{sqrt}(MS_{SEQ})$$

$$MS_{PIC} = \frac{\sum^{pic} (x - \hat{x})^2}{\text{no. of pels (= 101376)}}$$

where x: decoded pel value,
$\hat{x}$: original pel value

$$MS_{SEQ} = \frac{\sum^{seq} MS_{PIC}}{\text{no. of pics}}$$

2) SNR

$$SNR_{PIC} = 20\ LOG_{10}\ \frac{255}{RMS_{PIC}}$$

$$SNR_{SEQ} = 20\ LOG_{10}\ \frac{255}{RMS_{SEQ}}$$

3) Mean value of the step size

$$STEP_{PIC} = \frac{\sum^{pic} \text{step size of each GOB}}{\text{no. of GOBs (= 18)}}$$

$$STEP_{SEQ} = \frac{\sum^{seq} STEP_{PIC}}{\text{no. of pics}}$$

4) Mean value of the number of non-zero coefficients

$$NONZER_{PIC} = \frac{\text{no. of non-zeroes in picture}}{\text{no. of encoded blocks in picture}}$$

$$NONZER_{SEQ} = \frac{\sum^{seq} NONZER_{PIC}}{\text{no. of pics}}$$

5) Mean value of the number of zeroes before the last non-zero

$$ZER_{PIC} = \frac{\text{no. of zeroes before the last non-zero in picture}}{\text{no. of encoded blocks in picture}}$$

$$ZER_{SEQ} = \frac{\sum^{seq} ZER_{PIC}}{\text{no. of pics}}$$

6) Block type of Y

$$BTY_{PIC} = \text{no. of specified blocks in picture}$$

$$BTY_{SEQ} = \frac{\sum^{seq} BTY_{PIC}}{\text{no. of pics}}$$

7) Block type of C

$$BTC_{PIC} = \text{no. of specified blocks in picture}$$

$$BTC_{SEQ} = \frac{\sum^{seq} BTC_{PIC}}{\text{no. of pics}}$$

8) Number of bits

$$NOB_{PIC} = \text{no. of bits for specified information in picture}$$

$$NOB_{SEQ} = \frac{\sum^{seq} NOB_{PIC}}{\text{no. of pics}}$$