ISO / IEC JTC1 / SC29 / WG11
MPEG 92 /
Singapore, Januari 1992

CCITT SGXV
Working Party XV/1
Experts Group for ATM Video Coding

Doc AVC - 179
Singapore, Januari 1992

Source : CCITT SG XV - AVC
Title : DEFINITION OF END TO END DELAY
Purpose : Proposal

1. Introduction

This document describes all possible delays in a general way so that it can be applied to all possible coding schemes. The aim is to allow a proper comparison between coding schemes and to give reasonable grounds for determing the delay budget among all the elements that constitue end to end delay.

An ATM network environment is assumed, because this will be one of the most widely used media for conversational applications, and the conversational applications are the area in which end to end delay requirement should be most strictly met.

For the applications that do not utilize an ATM network the ATM related part of the delay can simply be ignored, and more delay could be assinged to other parts of the delay elements such as source encoding/decoding.

If the major concern is the comparison of coding schemes, the transmission delay can be excluded and the encoder and decoder can be placed back-to-back.

When the delay however should be calculated in order to meet some service requirements (e.g. low end to end delay) the transmission delay may not be neglected.

2. Description of delay

End to end delay of the system is defined as time elapsed between the camera output of a particular part of the picture and the monitor input of the corresponding part, while *encoding decoding delay* is defined as time elapsed between the encoder input and the decoder output when the encoder and decoder are connected back-to-back.

The different delays in the encoder are described in Figure 1(a), transmission delays in Figure 1(b) and the different delays in the decoder in Figure 1(c). For most blocks at least one example of possible delay is given in the following explanations. The delays described in Figure 1 should be applicable to all kinds of coding schemes :

• field based
• frame based
• schemes using B-frames

For clarification an example of delay calculation for all three coding shemes is given in section 3.

2.1 Description of the encoder delay

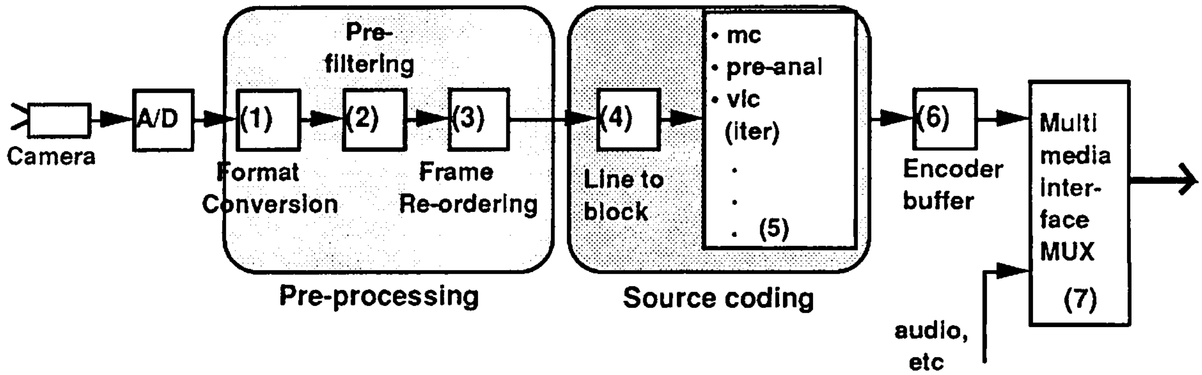

Figure 1(a) Encoder delay

The encoder delay consists of the following contributing parts :

- Format conversion (1) : e.g. • field ---> frame
  • spatial conversion
  • temporal conversion
- Prefiltering (2) e.g. temporal noise reduction
- Frame reordering (3) : e.g M = 3, see Figure 2
  I1 B2 B3 P4 B5 ... ---> I1 P4 B2 B3 P7 ...
- Line to block conversion (4) e.g. initial delay before processing
- Coding delay (5) e.g. • initial delay (search area MC)
  • buffer regulation ---> pre-analysis for rate control (e.g. multipass process)
- Encoder buffer (6) Note : Delay is not constant
- Multimedia interface (7) e.g. • synchronisation with audio (e.g. time stamps)
  • error correction
  Note : Delay is probably not constant

2.2 Transmission delay (8 ---> 12)

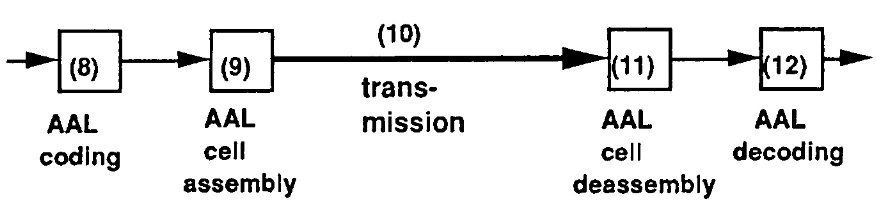

Figure 1(b) Transmission delay

The transmission delay can be excluded when comparing coding schemes.

The transmission delay is highly depending on the transmission media which is used. It may be typically 30 ms when optical fiber is used, but is 260 ms in one way delay when satellite transmission is used. For more clarification see AVC-85.

2.3 Description of the decoder delay

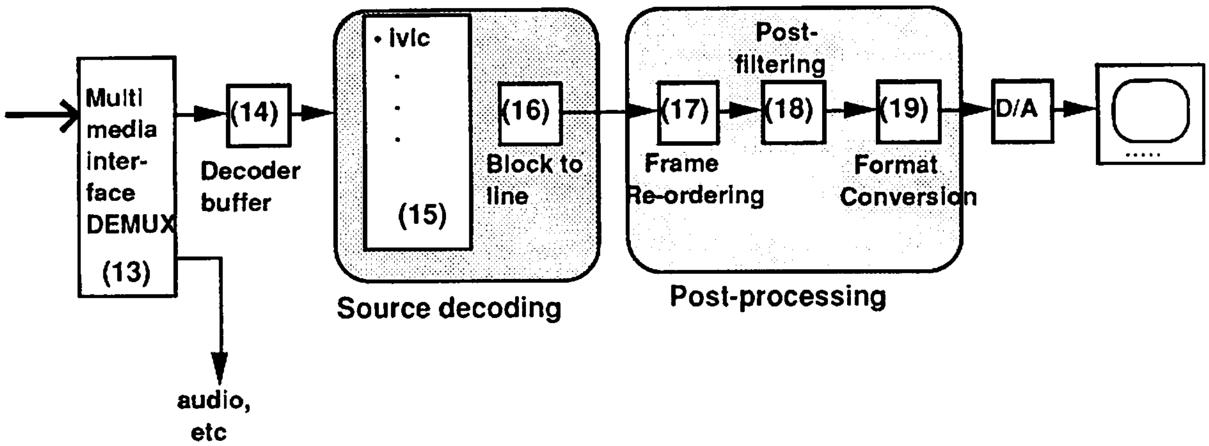

Figure 1(c)   Decoder delay

For the decoder - in simularity with the encoder - the following contributing parts should be encountered :

• Multmedia interface demux(13)	e.g error correction
• Decoder buffer (14)	Note : Delay is not constant
• Decoding delay (15)	e.g. inverse transform, if relevant inverse quantization
• Block to line conversion (16)	
• Frame reordering (17)	See also Figure 2.
• Post-filtering (18)	
• Format conversion (19)	e.g. • temporal conversion • spatial conversion • frame ---> field

2.4 Additional comments on delay in encoder and decoder

• Format conversion

Format conversion at an encoder from either 525 or 625 systems to some yet undefined common picture format, and the reverse process at the decoder may be required for conversational applications as a means to realize inter-regonial interworkability. (Example : An encoder using NTSC system and a decoder using PAL/SECAM system.)

If different coding formats will be allowed, a format conversion in the decoder might be required.

• Encoder and decoder buffer

Neither the encoder, nor the decoder buffer delay is constant. However as far as CBR coding is concerned and no frame dropping is used, the sum of the encoder and decoder buffer delay is constant. The following formula can be applied under the assumption of constant frame rate and constant bitrate :

$$D_{encoder\_buffer}\,(6) + D_{decoder\_buffer}\,(11) = cte = \frac{buffersize}{number\ bits/s}$$

- **Multimedia multiplexing delay**

The actual value of the delay incurred at this part varies largely with the multiplex scheme employed. Some examples are

- cell multiplex (VCI approach)
- SAR multiplex (packet approach)
- User multiplex (H.221 approach) see Fig. 1(a) block 7 and Fig. 1(c) block 13

See also AVC-129.

Although the amount of delay is independent of the coding scheme employed, we must note that in some cases a considerable amount must be allocated for this element to realize more friendly user interface and / or better interworkability with existing terminals.

3. Examples

3.1 Field based coding scheme

• Format conversion	$D_{enc}(1) = 0$ + possible spatial or temp conversion
• Prefiltering	$D_{enc}(2) = ...$
• Frame reordering	$D_{enc}(3) = 0$
• Line to block conversion	$D_{enc}(4) = ...$
• Coding delay	$D_{enc}(5) = ...$
• Buffer delay (enc + dec)	$D_{enc}(6) + D_{dec}(14) = \frac{\text{buffersize}}{\text{number bits/s}}$
• Multimedia +transm delay	$D_{enc}(7) + D_{transmission}(8\text{-}12) + D_{dec}(13) = ...$
• Decoder delay	$D_{dec}(15) = ...$
• Block to line conversion	$D_{dec}(16) = ...$
• Frame reordering	$D_{dec}(17) = 0$
• Post filtering	$D_{dec}(18) = ...$
• Format conversion	$D_{dec}(19) = 0$ + possible spatial + temp conversion

3.2 Frame based coding scheme

• Format conversion	$D_{enc}(1) = 1$ field delay + possible spatial or temp conversion
• Prefiltering	$D_{enc}(2) = ...$
• Frame reordering	$D_{enc}(3) = 0$
• Line to block conversion	$D_{enc}(4) = ...$
• Coding delay	$D_{enc}(5) = ...$
• Buffer delay (enc + dec)	$D_{enc}(6) + D_{dec}(14) = \frac{\text{buffersize}}{\text{number bits/s}}$
• Multimedia +transm delay	$D_{enc}(7) + D_{transmission}(8\text{-}12) + D_{dec}(13) = ...$
• Decoder delay	$D_{dec}(15) = ...$
• Block to line conversion	$D_{dec}(16) = ...$
• Frame reordering	$D_{dec}(17) = 0$
• Post filtering	$D_{dec}(18) = ...$
• Format conversion	$D_{dec}(19) = 1$ field delay + possible spatial + temp conversion

3.3 Coding scheme using B frames

In Figure 2 the problem of frame re-ordering is addressed. It can be seen that neither the encoder delay $D_{enc}(3)$, nor the decoder delay $D_{dec}(17)$ is constant. The total delay can however be expressed by the following formula :

$D_{enc}(3) + D_{dec}(17) = M$ frames — if $M > 1$ (B frames are used)
$D_{enc}(3) + D_{dec}(17) = 0$ — if $M = 1$, only forward prediction

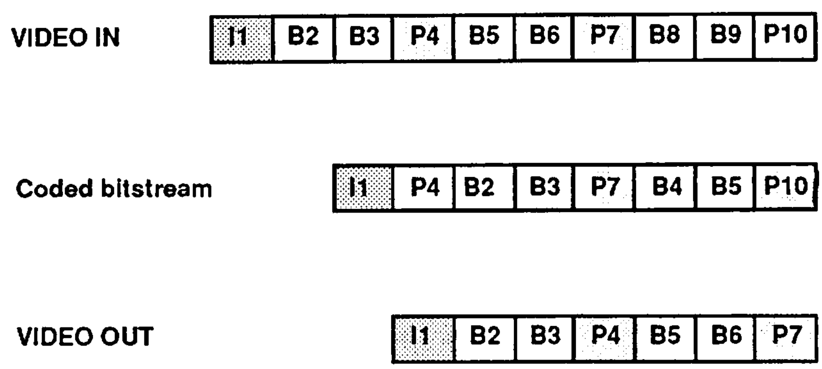

Figure 2 Frame re-ordering

- Format conversion $D_{enc}(1)$ = possible field to frame conv + spatial or temp conversion
- Prefiltering $D_{enc}(2) = ...$
- Frame reordering enc + dec $D_{enc}(3) + D_{dec}(17) = M$ frames
- Line to block conversion $D_{enc}(4) = ...$
- Coding delay $D_{enc}(5) = ...$
- Buffer delay (enc + dec) $D_{enc}(6) + D_{dec}(14) = \frac{\text{buffersize}}{\text{number bits/s}}$
- Multimedia +transm delay $D_{enc}(7) + D_{transmission}(8\text{-}12) + D_{dec}(13) =$
- Decoder delay $D_{dec}(15) = ...$
- Block to line conversion $D_{dec}(16) = ...$
- Post filtering $D_{dec}(18) = ...$
- Format conversion $D_{dec}(19)$ = possible frame to field conv + spatial + temp conversion

4. Conclusion

The definition of the end to end delay is given in general terms. It was found that the major contributing elements to the delay come from all the parts constituting the system, namely

- pre- and post-processing :
  - frame reordering (!!! B-frames)
  - format conversion
- source coding + buffer delay
- transmission coding and decoding

To realize a system that is most appealing to users in a sense of both picture quality and humane interface, it is necessary to find the best allocation of the delay among these three parts.

It should be noted that if pre-analysis is used in the encoder for better rate control, singnificant delay may be added because of its multipass configuration. This should also be considered when comparing the coding schemes.

It is appropriate to express the delay *both in number of frames (fields) and in ms*. This makes easy comparison possible and allows even to easily compare 50 and 60 Hz schemes.