CCITT SG XV
Working Party XV/1
Specialist Group on Coding for Visual Telephony

DOCUMENT #287
January 1988

Source : NTT, KDD, NEC and FUJITSU

Title : PROPOSAL AND CONSIDERATION OF BUFFER SIZE

---

1. Introduction

Investigations cocerning the relation between buffer memory size and picture quality have been done using Flexible Hardware equipment on the basis of initial compatibility check parameters. These resultant pictures will be presented at the meeting by VCR. Results show when buffer size is 64 Kbit, there is considerable deterioration of picture quality such as "waist cuts"(Note 1) in scene changes and during violent motion. If however coarse quantising parameters are forcibly selected to avoid the "waist cut", the picture degenerates considerably, especially after scene changes. On the other hand it was found that videoconferencing signals could be coded without problem if the buffer size was 128 Kbit or more. Based on these experimental results, we would like to make proposals regarding buffer memory size.

On the other hand the buffer size affects the coding delay. In buffer control where the total storage of the send/receive buffer is kept constant, the delay is about 31 msec per 10 Kbit, it means 400 msec at 128 Kbit buffer, and this may be not permissible. For reference purposes, we shall mention here a report describing a control method where even if buffer size is increased, the delay time caused by the buffer memory can be decreased if the receiver side buffer is read rapidly. And some examinations on the fluctuation of delay time caused by this control are introduced. And we also propose a method to specify the delay caused by buffer memory and its value.

2. Structure of Buffer Memory

The structure of the buffer memory may be arranged as (1) or (2) in Fig. 1. Arrangement (1) is the buffer

Note 1 : If the transmission buffer becomes full in the middle of picture, there occures discontinuity between upper and lower parts.

structure assumed in the reference model of the simulation works, but as we shall show later, it can be controlled in frame units.

Arrangement (2) was introduced by Japan at the Montreal meeting (Doc. #116). As there is an input buffer the transmission buffer is considered to be infinite, so there is no possibility of overflow and no "waist cuts" occurrence. The coder repeats start and stop in GOB units.

According to either one of the above arrangements being chosen, the required transmission buffer memory size will change. Judging from the discussions in the meeting so far and the way in which the Flexible Hardware equipment is composed, however we assume arrangement (1) as a common base of the consideration of the buffer size quantity.

Even in this arrangement two types of control structures are possible, such as, a decoder is operated so as to maintain TB + RB = constant, and a variable delay structure where the source decoder is operated each time full coded bits for one frame have been stored in RB (Montreal meeting, Doc. #111). We shall discuss the later control structure from the viewpoints of delay time, sound and picture synchronization, and compatibility.

3. Buffer Capacity, Buffer Control Method and Delay Time

Since one type of Japanese hardware has a 256 Kbit buffer memory, there will be delay of 800 msec (256/320) if we adopt the constant delay system. However this codec adopts a variable delay system wherein the codes for one frame are decoded after they are fully stored in the receiving buffer, so that the delay can be decreased. It has been confirmed through decoding the DIS test data that there is no problem in compatibility even if the variable delay method is adopted.

(1) Delay due to asynchronization of frame pulses

It has been suggested that with the variable delay system, delay may occur due not only to buffering state on the coder side, but also the use of asynchronous frame pulses on the decoder side. As a result, even if the coder is operating at a constant rate (2 : 1) as shown in Fig 2, the frame rate on the decoder side may vary within 1 frame duration if the delay time on the coder side (buffer state) varies.(Note 2) It is possible therefore that this could lead to some visual unnaturalness in the picture. When we carried out simulation tests to determine the extent of this effect, we found it was less annoying than might have been expected.

Note 2 : Here is assumed an architecture that the source decoder operates only when two or more PSCs except those of dropped pictures are inside the receiving buffer.

(2) Delay time inherent in the variable delay system

As there is delay time of 1 or 2 frames in TB, 0 or 1 frame in RB and 2 frames in the code volume (in case of 2 : 1 frame dropping operation), a total delay of approx. 4 or 5 frames can be envisaged. In practice, however, TB is adjusted to 256 Kbit in order to accomodate increase of code volume during scene cuts, and so this delay time could increase at a glance. In the constant delay system, there is a fixed delay of 6 frames (64 Kbit) to 24 frames (256 Kbit).

(3)Lip Sync (sound and picture synchronization)

As we are using the variable delay system where the coded frame rate is variable, it is difficult to achieve exact lip sync, however if the buffer data is controlled around any given point, some degree of lip sync can be attained as shown in Fig. 3. A method for determing the audio delay could be conceived for example by caluculating the average value of the buffer state from Fig. 4 or a similar method.

4. Conclusion

It is clear from the above that even if the buffer memory is made relatively large, the magnitude of the coding delay due to the buffer memory can be decreased. We therefore propose from the viewpoints of improving coded picture quality and convenience of use for videoconferencing, to insert the following sentences in Draft Recommendation H.12X, Section 5.

> 5. Video Data Buffering
>
> The buffer size on the sending side shall be 128 Kbit (Note). The overall delay time on the sending and receiving sides due to this buffer memory, shall however be controlled such that it does not exceed 132 msec when every picture (29.97 frame/sec) is coded.
>
> Note : K = 1024. This size is applied to n = 1 for n * 384 kbit/s. For higher values of n, further study is needed.

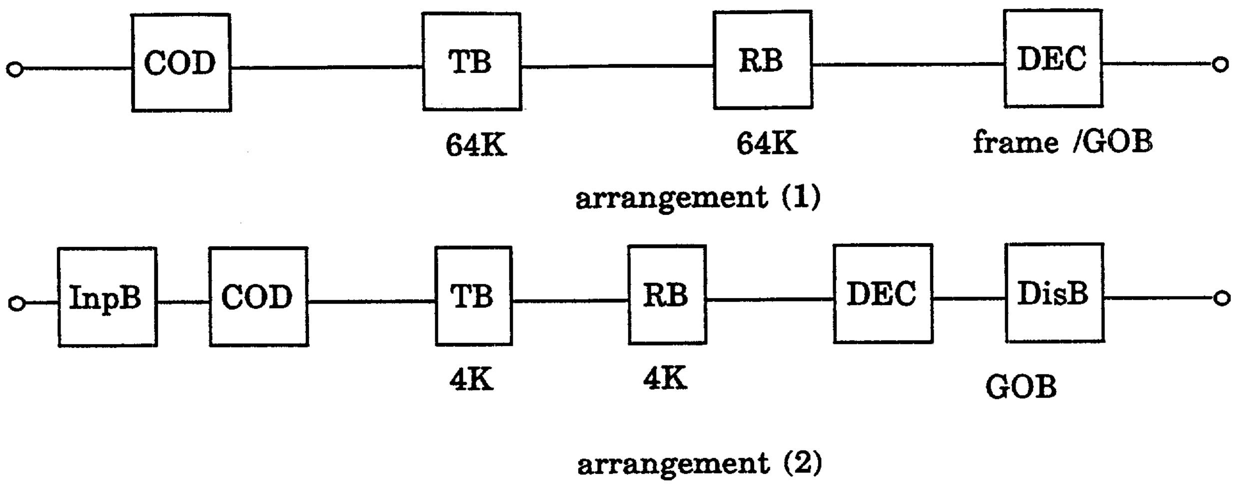

Fig. 1 Structure of buffer memory

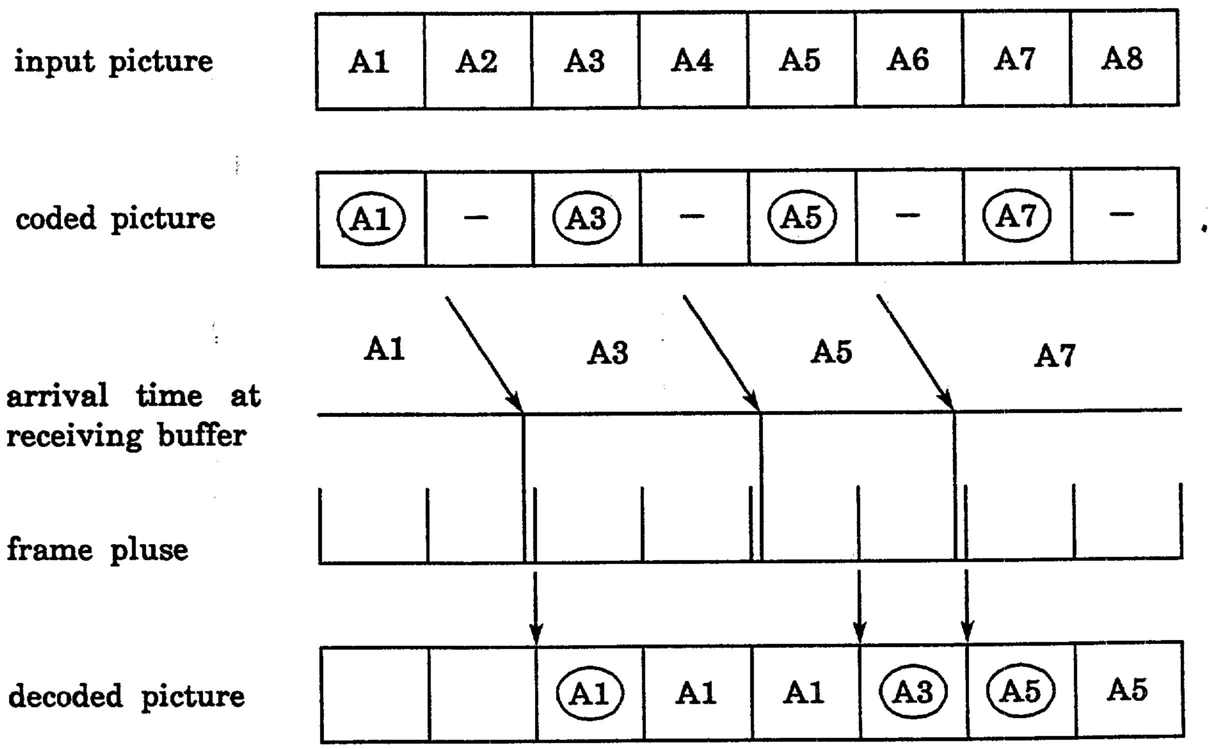

Fig.2 Variation of delay time due to asynchronization of frame pulses

input video	1	2	3	4	5	6	7	8	9	10	11
coded video	①	–	③	–	–	⑥	–	⑧	⑨	–	⑪
decoded video	①	1	1	③	3	⑥	6	⑧	8	⑨	⑪
decoded audio	0	1	2	3	4	5	6	7	8	9	10

Fig.3 Time relation between picture and sound

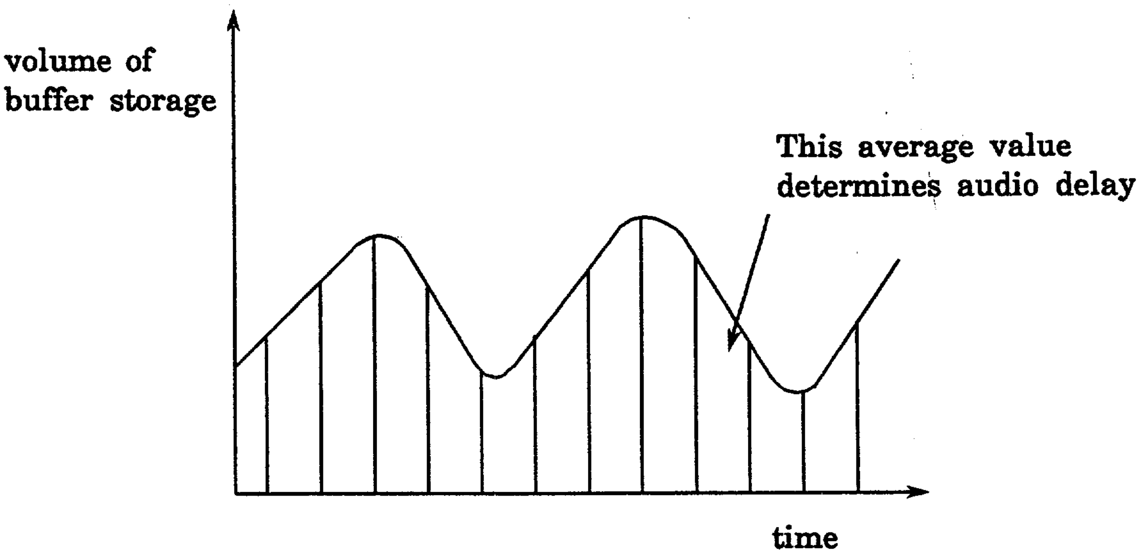

Fig.4 Determination of audio delay