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

Doc #203
March 1987

Title: Audio Delay in n*384kbit/s Codec

Source: UK

The n*384kbit/s video coding algorithm introduces significant delay in the video path. A compensating delay in the audio chain may be necessary. The delay could be introduced in three ways:

1. All at the encoder
2. All at the decoder
3. Split between the encoder and decoder

The first two methods may be slightly more economical in hardware than the third but suffer from the disadvantage that the video delay at the codec in the other location must be standardised. Hence, the third method is proposed as this leaves more scope for choice of video processing and audio delay matched to it.

A CCIR report [1] addresses the question of permissible delays between sound and vision. With critical material:

'For sound delayed with respect to vision, 140 ms will produce, approximately, a "just perceptible impairment" for 50% of the observers.'

'For sound advanced with respect to vision, 70 ms will produce, approximately, a "just perceptible impairment" for 50% of the observers.'

It is suggested, therefore, that the specification provisionally call for the relative delay between encoded audio and encoded video at the channel output of the n*384kbit/s coder to be within half of the above figures when the encoder buffer is empty. To achieve this it may be necessary to change the padding delays inserted at the coder and decoder for different bit-rates in the range 384kbit/s to 2048kbit/s. However, the audio delay should not be altered in line with any dynamic variation of picture temporal subsampling ratio.

There is a human factors requirement in services with bi-directional sound to minimise the total round trip delay time. This may be in conflict with the objective of maintaining simultaneity between sound and vision. Also, at the lower picture rates likely to be imposed by the video coding technique at small values of 'n' it is likely that the above limits can be relaxed.

Further study would thus be beneficial before a final specification is adopted.

Reference [1]. CCIR Section CMTT E: Report 412-3. Transmission Time Differences between the Sound and Vision Components of a Television Signal.