Proceedings of the Institute of Acoustics

MAINTAINING SOUND AND PICTURE SYNCHRONISATION IN A COMPLEX TELEVISION PRODUCTION ENVIRONMENT

D.G. Kirby.

British Broadcasting Corporation, Research and Development Department, Kingswood Warren, Tadworth, Surrey. KT20 6NP. England.

ABSTRACT.

As the sophistication of video and audio equipment increases, so, almost inevitably, does the signal delay introduced by them. These delays rapidly accumulate and lead to a distracting loss of synchronisation between sound and picture, (loss of "lip-sync"). This is becoming a widespread problem which can affect all types of television output, including both live and pre-recorded programmes and films. Piece-meal solutions to this problem have used isolated audio delays to compensate for delays in the video but this approach can only offer a partial solution.

An alternative approach is described in this paper which has been developed in conjunction with BBC Studio Operations, Television. In this approach the delay to both video and audio signals is allowed to accumulate as they pass through an area, but a numeric code is added to the video signal signifying the total differential delay encountered. At strategic points, such as a studio output, the accumulated error is corrected to re-establish synchronisation.

The system is currently undergoing field trials at the BBC's Television Centre and the proposal has been submitted to the EBU for further discussions with a view to standardising this approach.

1. INTRODUCTION.

The increasing use of television equipment which introduces a significant delay into the video or audio signal is leading to a serious problem in maintaining the correct synchronisation of these signals. The widespread use of video synchronisers, sometimes in cascade, is further compounding the problem, particularly as these devices introduce a varying video delay.

Currently the problem of maintaining correct audio synchronisation is addressed in one of three ways, either:

1) apply no correction,
2) apply a fixed correction, or
3) use a tracking "A/V sync" audio delay.

The first approach is probably used more frequently than it should be and soon leads to a significant accumulation of small errors which may individually be acceptable. The second solution can be used where a fixed signal route is in use and its delay can be measured or, alternatively, a variety of signal delays may occur, as routings change, and a fixed delay offers a compromise correction. The third uses an audio delay which automatically tracks the difference in the timing of the input and output video signals across an item of video equipment, i.e. the difference in timing of the video is measured and applied to the audio. This is frequently used with video synchronisers where the delay required will change as different remote sources are selected.

Clearly the first approach is inadequate; the other two require the expense of installing many audio delays and will lead to a degradation of the audio quality owing to the cascaded conversions in the delays. The third method additionally suffers from the practical difficulty that both audio and video signals must be brought together at the tracking delay. As installations frequently have video and audio in separate technical areas, this presents a significant additional complication.

Figure 1 shows the typical studio arrangement used at present with a tracking delay correcting the audio

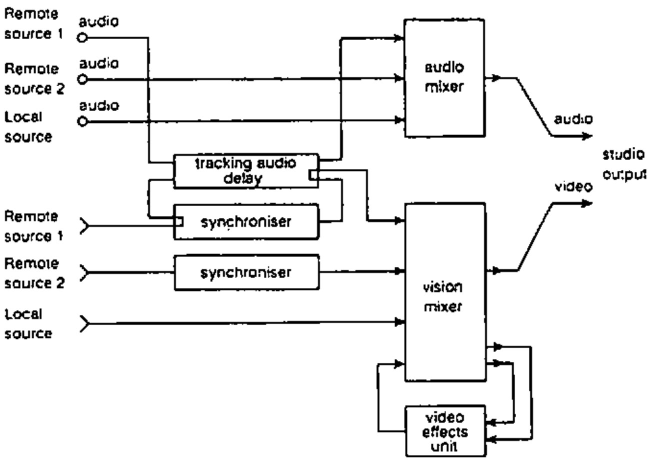

Figure 1. A typical audio and video configuration within a studio.

for Remote Source 1 and no correction applied for Remote Source 2.
The digital video effects unit shown around the vision mixer causes yet a further problem as these generally introduce a further 20 or 40 ms delay into the video which cannot be corrected for as the DVE is not always is circuit but is switched in and out as required. This gives rise to the situation where, as the DVE is selected a few seconds before an effect is used, the audio goes out of synchronisation.

MAINTAINING SOUND AND PICTURE SYNCHRONISATION IN TV PRODUCTION

2. THE NEW PROPOSAL

An alternative solution to the problem of maintaining synchronisation is to carry an indication of audio/video mis-timing within the video signal itself. This indication can be carried conveniently on a spare line of the vertical blanking interval (VBI). A video signal which is correctly timed with its audio carries either no code or a code indicating zero offset. As the video signal passes through a device which introduces significant delay, the code is modified to reflect the additional delay which is being incurred. Thus as a video signal passes through a production area, the delay code carried in its VBI is modified, each time an additional delay is encountered.

At points where a significant audio delay is introduced, for example by audio coding equipment, the accumulated delay code must be reduced by the corresponding amount. In cases where the audio is delayed more than the video this will result in a negative delay code being carried in the VBI. An alternative implementation where AES/EBU digital audio signals are being used could carry a similar audio delay code in the user bits. This would avoid the need to modify the code in the VBI as audio delays are encountered and would give a separate indication of the accumulated audio delay.

At any point in the signal path, the delay code in the VBI (or the difference between video and audio delay codes), represents the total mis-timing between the video and audio incurred. At significant points in the area, such as the studio output, the delay code can be read and used to set an audio delay to re-establish correct synchronisation. Once this correction has been applied the delay code is set to zero or blanked from the signal.

As the delay code will always reflect the delay incurred, changes of signal source or route are of no consequence; the correcting audio delay at the output will follow changes in the accumulated delay as reflected by the inserted delay code.

3. IMPLEMENTATION

As many video processing systems only pass the active picture area, it is necessary to read the delay code at the input to such equipment and insert the modified code as the video emerges from it. Although this function would preferably be built into new video hardware, existing equipment will require an external unit to read and subsequently re-insert the delay code.

An example of the use of this system is shown in figure 2, which shows two video synchronisers supplying signals to a vision mixer which also has a video effects unit associated with it. The signals from the remote sources first loop through their respective delay code inserters which read any delay code already present on the signals.[1]

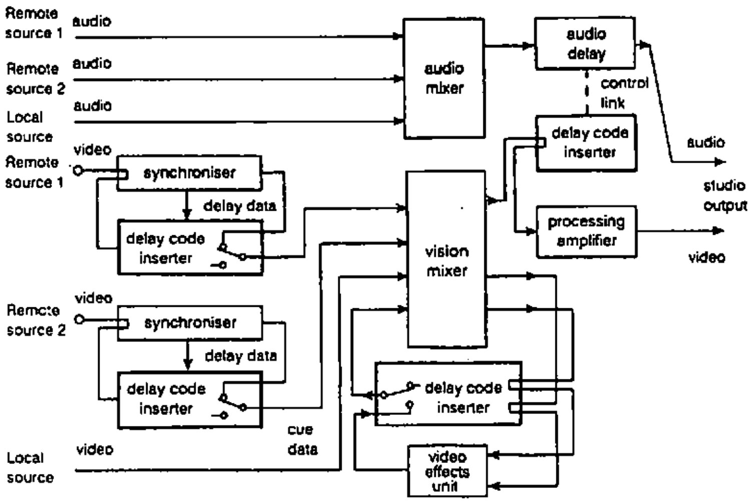

Figure 2. A typical studio set-up with automatic audio-video synchronisation

The delay introduced by each synchroniser in re-timing its video is signalled to the delay code inserter via a data connection. The delay code inserter calculates the accumulated delay value and inserts the appropriate data into the VBI of the synchroniser's output video.

When either remote source is selected it will appear on the output of the vision mixer and loop through the input side of the delay code inserter which is connected to the mixer output. This unit reads the delay value and controls the compensating audio delay, which changes its delay inaudibly, as required. The video finally passes through the normal processing amplifier which will remove the inserted data. When the video signal is routed through the video effects unit, the delay code inserter associated with that unit will modify the code to reflect the additional two fields of delay. In this way the delay code always indicates the accumulated delay although signal routings and equipment delays may themselves be changing.

Normally such a source would not contain a delay code as its timing would have been corrected. However, there are some operational circumstances (for reverse feeds, etc.), where minimum delay is required and it is better to correct the mistiming of a remote source within the studio. In this case the signal would contain a delay code.

4. Data Requirements

4.1 Range and Resolution

The maximum uncorrected delay to be encountered is unlikely to be significantly greater than 500ms and certainly less than 1 second. It is possible that small negative delays may need to be transmitted if time-consuming audio processing takes place but the accumulated value along the chain would have to be positive to be useful - unless the reader is connected to a video delay.

CCIR Recommendation 1042-E gives maximum acceptable differences of 20ms (sound ahead of vision) and 40ms (vision ahead of sound), for "international exchange of programmes". This is a minimum resolution requirement.

However, if the delay encountered by a data bridge is not an exact multiple of transmitted units, a rounding or truncation error will be introduced. When several data bridges are cascaded, the errors will accumulate, until, in the worst case, they may exceed the figures given above.

Discussions indicate that a typical programme chain will contain three vision delaying devices. In order to give room for manoeuvre, it is therefore sensible that the system should allow five bridges to be cascaded.[2] In this case a delay resolution of 5ms is required and this would allow delays of -100ms to +535ms to be represented in seven bits. However, if seven bits are needed, it is no more difficult to use a whole byte (8 bits). This enables the resolution to be increased to 4ms (10 bridges maximum) while also giving a greater range of delays: -128 to +892ms is a sensible range from the point of view of the binary logic.

It is therefore proposed that the delay should be represented by a byte of data covering the range -128 to +892ms in 4ms steps.

5. Format of Inserted Data

5.1 General requirements

Given the relatively low data rate required a simple scheme of conveying the data is preferred. The method adopted must be suitable for use in analogue and digital video environments and should preferably avoid the need for transcoding equipment when changing between these standards.

Several existing data formats have been considered: teletext, International Insertion Data [1], and the data

[2] Other situations, such as complex international contributions, may exceed these original design figures. Discussions are in progress to determine more suitable figures for these new applications.

areas of vertical interval time code (VITC) [2] and insertion test signals (ITS). The teletext and the International Insertion Data formats are unnecessarily complex for this application and would require equipment to recode the data when changing between analogue and digital standards. The data fields within VITC and ITS are inconvenient to use and would require either a correct VITC or ITS to be inserted neither of which is appropriate and could be misleading.

5.2 Data coding

As there is no need to make the data dc-free, a simple "non return to zero" (NRZ) code will suffice. Black level represents logic zero, while a voltage near to peak white, to give the best data integrity, represents logic 1. In addition, a 'start' pulse to synchronise the data receiver avoids the need to ensure critical positioning of the data on the video line and an address preamble identifies the data type and allows for future expansion.

In order not to effect the picture, the data needs to be placed on a spare line during the vertical blanking interval (VBI). The line number is not important, and will depend on which lines are available.

The proposed structure for the data is shown in Figure 3.

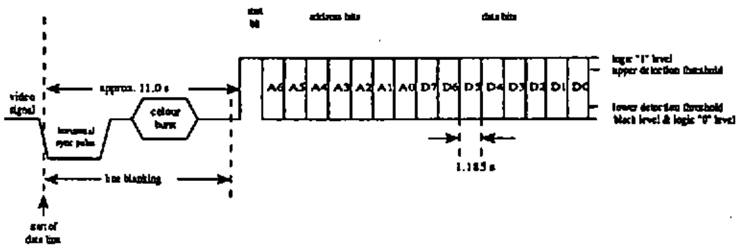

Figure 3. The format of the inserted data

5.4 Data Rate

Before teletext is added to a broadcast signal, there is no undue pressure on the use of the VBI, so there is no strong requirement to minimise the amount of the television signal taken up by the data. Indeed, the lower the data rate, the more reliably it can be transmitted. However, it is complicated and wasteful for the data to be spread over more than one line in each field. Therefore, the data should be accommodated on one active line of 52μs. This figure is halved if HDTV is taken into consideration. For the 16-bit block, the resulting minimum bit rate is 0.64Mbits/sec.

The bit rate is arbitrary for an analogue signal, but, in the digital domain, hardware is simplified if each bit lasts for an integral number of sampling periods. Unfortunately, the lowest common multiple between the sample rates of 12.5MHz for component, and 17.73447MHz for composite digital video, gives a bit rate less than the minimum specified above. Therefore, a compromise has to be found and the best match is between 16 component samples and 21 composite samples, corresponding to data rates of 0.84375 and 0.844499Mbit/s respectively. The average rate of 0.844124Mbit/s should ideally be used in the analogue domain (although in practice 844kbit/s would be acceptable). As HDTV is sampled at a multiple of 13.5MHz, the proposed data rate will be equally applicable in this case.

6. Operational Considerations

Video mixers do not introduce significant delay and are assumed to pass VBI from lines 10 and 323 onwards, if their associated processing amplifier is bypassed or reconfigured[3]. It is worth doing this in view of the number of inputs which would otherwise have to be connected to delay code inserters. Data will pass through mixers unhindered except when a fade or a wipe is taking place. If a partial fade or wipe is held, a corrupted data value may be repeated enough to be interpreted as a correct delay.

A held fade produces codes which are the sum of the weighted bit levels of the data from the sources. The voltages will therefore be in between the two normally expected, and can be detected by a window comparator. This inevitably reduces the noise immunity of the system, but the low noise video path should not lead to many correct words being interpreted as erroneous.

7. Use with Digital Video

The CCIR Rec 656 digital video standard [3] includes provision for the inclusion of ancillary data and this facility could be used to convey the proposed delay data.

However the use of the ancillary data capacity in this way will complicate the interfaces between analogue and digital video standards. Transcoding equipment will be required between each standard leading to an overall increase in system cost. If video ADCs and DACs pass the VBI the proposed insertion of data will be carried between standards without any additional equipment. It is therefore preferred to retain the simple data insertion technique proposed for digital video applications where possible. It is recognised, however, that some digital video equipment will probably not pass signals outside the active picture area. In this case transcoding equipment to place the delay code in the ancillary data fields will be necessary.

---

3 This assumption, whilst true for many mixers, requires a wider discussion with manufacturers.

MAINTAINING SOUND AND PICTURE SYNCHRONISATION IN TV PRODUCTION

8. FUTURE DEVELOPMENTS

The system outlined in this document is currently undergoing field trials at the BBC's Television Centre. Initial results are encouraging and as a result the installation is to be moved to the News area to undergo a more rigorous trial. In anticipation of a successful outcome, this proposal has been put forward for consideration amongst broadcasters and equipment manufacturers. It is hoped that if a standard approach can be agreed then the system can be incorporated into new video and audio equipment, minimising the additional costs incurred and requiring the installation of only the minimum number of compensating audio delays.

9. ACKNOWLEDGEMENT

The author would like to thank the BBC for permission to publish this paper.

10. REFERENCES

1."Specification of insertion data signal equipment for International transmissions"; EBU Tech. 3217-E, third addition, June 1977.

2."EBU time-and-control codes for television tape-recordings (625-line television systems)", EBU Tech 3097-E, Third Edition April 1982.

3."Interfaces for digital component video signals in 525-line and 625-line television systems operating at the 4:2:2 level of recommendation 601". CCIR Recommendation 656, document number 11/BL/37-E. May 1992.