A new technique to maintain sound and picture synchronization

D.G. Kirby (BBC)

M.R. Marks (BBC)

It is becoming more common to see television programmes broadcast with the sound and pictures out of synchronization. Such timing errors can occur very easily today, due to video and audio signals being processed separately, for example by synchronizers and video effects units.

The BBC has been concerned with signal synchronization for some time and has explored various techniques to control it. The most promising approach uses delay codes, carried within the signals themselves, to indicate the amount of delay which each signal has experienced. These codes are updated whenever a signal is further delayed by equipment, and therefore they indicate the extent of the mistiming between sound and vision at any point in the production chain.

When required, typically at the point of recording or transmission, the accumulated delay values are read from the signals and used to apply a compensating delay to re-synchronize the sound and pictures.

1. Introduction

As the sophistication of video and audio equipment increases, so – almost inevitably – does the delay experienced by the signals passing through them. Unless great care is taken to match the delays in the audio and video paths, any small differences can rapidly accumulate and lead to a distracting loss of synchronization between sound and picture (most notably, loss of “lip-sync”). This is already becoming a widespread problem in television production and it affects all types of programmes, including live and pre-recorded material, and films. There are an increasing number of instances where programmes are broadcast with a noticeable error in synchronization; as low bit-rate audio and video coding systems become more widely available, so the scope for such timing errors will increase considerably.

At present, piece-meal solutions to this problem have been employed using isolated audio delays to compensate for delays in the video, but this

Original language: English.
Manuscript received 2/5/95.

approach can only offer a partial solution. For example, digital video effects units (DVEs) are normally switched in and out of circuit when required for the programme. For the time that the DVE is selected before the effect, to when it is deselected afterwards, the video is delayed by typically 40 ms and loses synchronization with the audio.

An alternative approach to control this problem, based on a proposal made by BBC Studio Operations, Television[1], is described here. In this approach, the delay to both the video and audio signals is allowed to accumulate as they each pass through an area, but a numeric code is added to the video signal (and possibly the audio signal, if digitally transmitted) to signify the total delay encountered. At strategic points, such as at the studio output or network control centre, the numeric codes are extracted from the signals to give the accumulated error at that point. An appropriate correction can then be applied to re-establish synchronization prior to recording or transmission.

Equipment has been developed to demonstrate this procedure and has been successfully used in the News Studio at the BBC's Television Centre in West London. Following the success of this field trial, additional studios and equipment are now being equipped with the system to give wider coverage and control of timing errors in the BBC's broadcast chain.

2. Current approaches

Currently the problem of maintaining correct audio synchronization is addressed in one of three ways:

- allowing the signals to pass uncorrected;
- applying a fixed correction;
- using 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 only be acceptable individually. The second method can be used for a fixed signal route where the delay can be measured or, alternatively, where a variety of signal delays may occur as routeings change, with a fixed delay offering a compromise correction. The third approach 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 synchronizers where the delay required will change with time and as different remote sources are selected.

Clearly the first approach is inadequate. The other two require the expense of installing many audio delays and the audio quality will degrade owing to the cascaded conversions in these delays. The third method additionally suffers from the practical difficulty that both the 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.

Fig. 1 shows the typical studio arrangement used at present. A tracking delay corrects the audio for remote source 1 but no correction is applied to remote source 2. A reverse cue feed is provided to remote source 1, but this is unsatisfactory because it is fed back after the tracking delay has been added, causing disturbance to commentators who are able to hear their own voices delayed.

The digital video effects unit causes another problem as it generally introduces a further 40 or 80 ms delay into the video chain. This delay cannot be corrected, as the DVE is not always in circuit but is switched in and out as required. Hence, whenever the DVE is selected, synchronization is lost.

Considering just this simple arrangement, it is clear that there is significant scope for unintentionally introducing sound-to-picture timing errors; in a fully equipped studio complex, the problem soon becomes almost impossible to control.

3. A new approach

Various ideas to control this situation have been considered; the most promising solution which was felt to be worthwhile investigating involves carrying an indication of audio/video mistiming within the video signal itself, e.g. within the vertical blanking interval (VBI). Where a video signal is correctly timed with its audio, it 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 the video and audio signals pass through a production area, the delay code is updated each time either signal encounters a significant additional delay to indicate the cumulative timing error up to that point.

Where a significant audio delay is introduced (for example, by audio coding equipment) the accumu-

[1] and in particular, Mr. Larry Goodson.

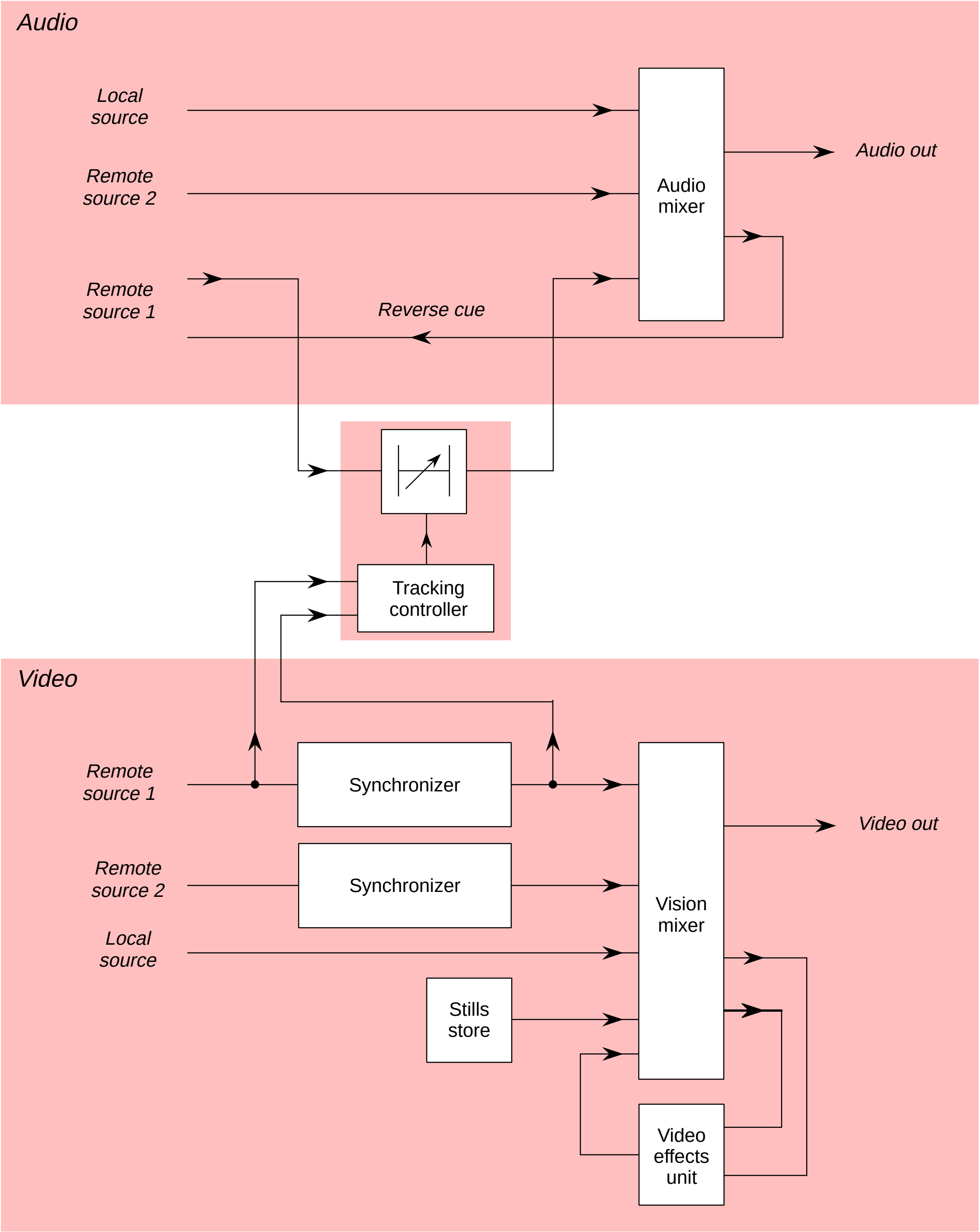


Figure 1
Typical audio and video configuration within a studio.

lated 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. An alternative implementation, where AES/EBU digital audio signals are being used, could carry a similar audio delay code in the embedded user bitstream. This would avoid the need to modify the code in the video signal as audio delays are encountered, and would give a separate indication of the accumulated audio delay.

At any point in the signal path, the video delay code (or the difference between video and audio delay codes) represents the current mistiming between the video and audio. 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 synchronization. This device must be capable of changing the delay inaudibly, and as rapidly as possible. Once the correction has been applied, the delay code is set to zero or blanked from the signal.

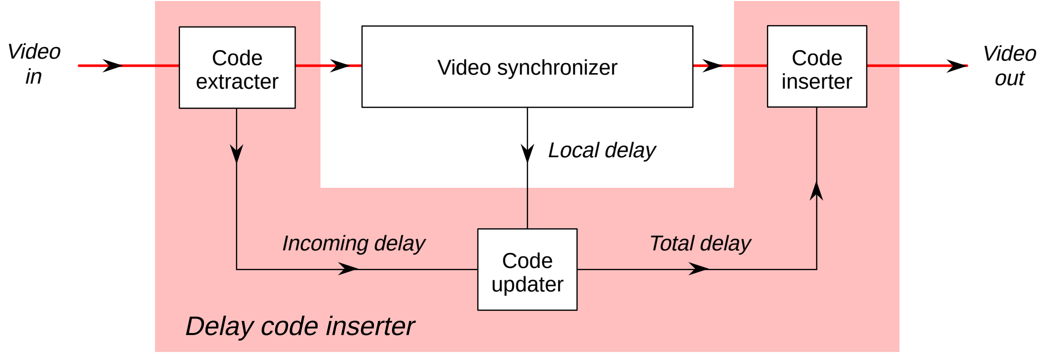


Figure 2
Video synchronizer with associated delay code inserter.

As the delay code will always reflect the accumulated delay that has occurred, changes of signal source or route are of no consequence; the audio delay required at the output to synchronize the sound and vision signals will be determined in real time by the inserted delay code. It should be noted that it is not necessary to insert codes on all signals, as the absence of a code implies that no delay has been encountered and hence no correction is required. Only signals which encounter a delay (and may therefore become mistimed with their corresponding audio or video signal), will require a delay code to be inserted.

4. Using the delay codes

4.1. Basic operation

Before considering the operation of the system in more detail, it may be clearer to consider how the delay codes are implemented in the simple case of a video synchronizer. *Fig. 2* shows the arrangement employed.

The delay code inserter first extracts any delay code information already present on the incoming video signal. The local delay introduced by the synchronizer itself is signalled directly to the delay code inserter by using a data connection, or similar arrangement. The total delay, i.e. the sum of the incoming and the local delay values, is re-inserted into the video signal as it passes from the synchronizer through the code inserter to the video output. The delay code which is inserted into the video signal will therefore always follow the delay introduced by the synchronizer.

It can be seen from this description that the delay code does not pass through the synchronizer itself; it is therefore not a requirement for the synchronizer to be transparent to the inserted data. (Other equipment such as the vision mixer, however, is required to pass the delay codes.)

4.2. Operation within a studio area

The operation of this simple arrangement within a basic studio is shown in *Fig. 3*. In this arrangement, two video synchronizers and a stills store supply signals to a vision mixer which also has a video effects unit associated with it. The two video synchronizers and the DVE are fitted with delay code inserters, as already described. For clarity, the delay code inserter equipment has been split into functional elements, labelled “A”, “B”, “C” and “D” in *Fig. 3*.

Tracing the route taken by remote source 1 illustrates the operation of the system. The incoming signal first enters the delay code extracter (labelled “A”) which extracts any delay information already present. The further delay introduced by the synchronizer in re-timing its video signal is indicated via a data connection to the code updater (“B”), which calculates the total delay and instructs the code inserter (“C”) to insert the appropriate data into the synchronizer’s output video. In addition, known timing errors, introduced perhaps by a contribution link, can be added to the code by the use of switches on the code inserter.

When this source is selected by the vision mixer it passes through to the studio output. However, before leaving the studio, the signal first passes through a delay code extracter (“A”) which extracts the accumulated delay information. This value, which now represents the total mistiming between the sound and vision signals, is passed to the audio delay device which then changes as rapidly as possible, but inaudibly, to the required compensating delay. Correct synchronization of sound and picture is then restored. The delay code must now be removed or set to zero so that further correction is not applied downstream;

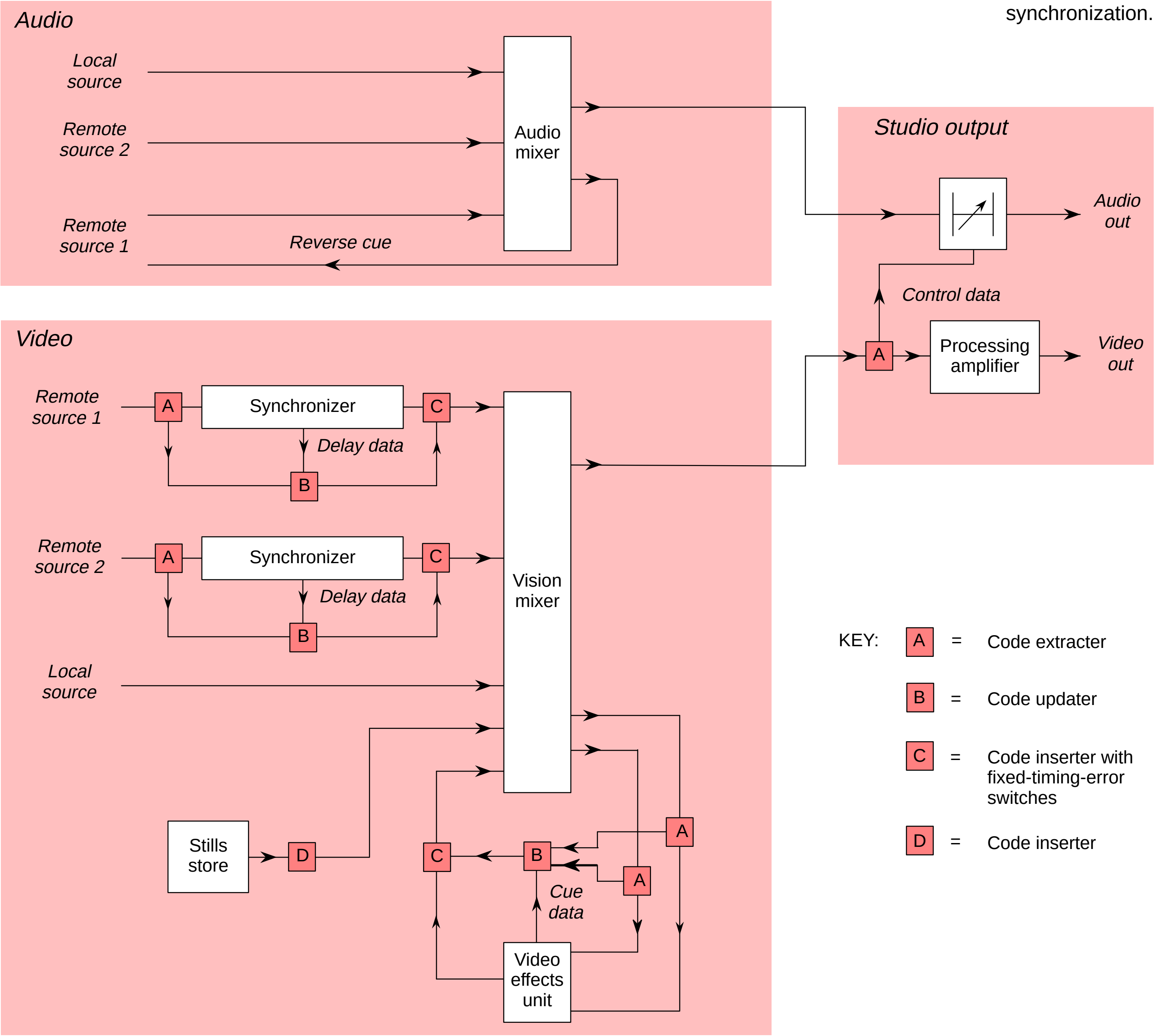
this is most conveniently done by the processing amplifier which is normally installed on studio outputs.

Whenever the selection on the vision mixer changes, the delay code reader at the studio output extracts the new delay code information from the video signal and introduces an appropriate audio delay to eliminate the mistiming errors present in the newly-selected vision source. If a correctly-timed source is selected, it will not contain a delay code. In this case the lack of code is detected and the compensating audio delay will change to zero, as no correction is required. (In passing, it should be noted that signals from CCD cameras in the studio could have a fixed code added to their outputs, to correct for the video delay inherent in their scanning process.)

4.3. *Operation with the DVE*

When the digital video effects unit is brought into use, its input video signals will first pass through the associated delay code extracter to read any existing delay information. The updating of the delay code will take place as described earlier but, in this case, it is slightly more complicated because multiple video sources may be contributing to the one DVE output. In order to calculate the most appropriate delay for the sources in use at that moment, the DVE unit supplies the code updater ("B") with cue data to indicate which input signals are in use.

Figure 3
Typical studio setup with automatic audio-video synchronization.



If only one video source is in use, then the delay code inserter will take codes from that signal. However, if more than one input to the vision mixer is selected for processing, the source with the greatest code value will be used, because the subjective effect of too great an audio delay is less severe than having the sound early. Alternatively, a weighted average of selected signals could be taken, or an intermediate delay value which would maintain all sources within an acceptable range.

In this way the delay code carried by the video signal will be updated appropriately as the DVE is switched in and out, and the compensating audio delay will follow the changes accordingly for the duration of the effect.

4.4. *Picture sources without sound*

A picture-only source, such as a stills store, can use a unique code value to signify that a delay to its signal is unimportant. This special code is passed unchanged through any subsequent delay code inserters, and has the effect of freezing the compensating audio delay at its previous value. In this way, whilst this picture source is in use, the audio delay does not unnecessarily reduce to zero, as would happen if there were no code on its signal.

The purpose of this refinement is to cover the situation where the vision cuts to another picture, for example a graphic, without interruption to the audio, e.g. a commentary. Freezing the audio delay whilst the graphic is shown avoids the inevitable momentary loss of synchronization that occurs when the previous picture source (the commentator) is re-selected and the compensating delay corrects itself back to the previous value.

5. *Format of the delay code*

5.1. *Range and resolution requirements*

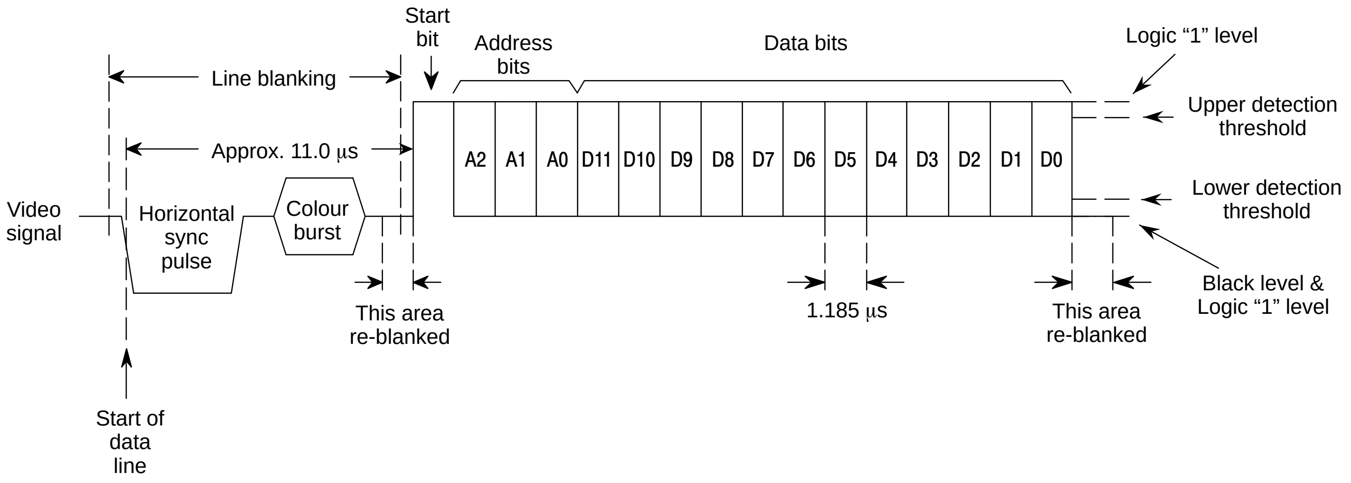
Given the relatively low data-rate required in this application, a simple scheme of conveying the codes is preferred. The method adopted must be suitable for use in analogue and digital video environments and should preferably avoid the need to provide transcoding equipment when changing between these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

The maximum uncorrected delay to be encountered is unlikely to be significantly greater than 500 ms, and will certainly be less than 1 second within a studio. However, in exceptional circumstances (such as in the case of some international contributions), larger values may be needed. It is also possible that small negative delays may need to be transmitted if time-consuming audio processing has taken place.

The basis of a resolution requirement for the “international exchange of programmes” is given in CCIR Recommendation 717 [1]; the maximum acceptable mismatch is specified as 20 ms for sound ahead of vision, and 40 ms for vision ahead of sound. However, to avoid cumulative rounding or truncation errors, a better resolution than this is needed. A resolution of 1 ms allows at least 20 code inserters to be cascaded before worst-case errors exceed 20 ms.

After due consideration, a 12-bit code was chosen to give a delay range of – 511 to + 3583 ms with a resolution of 1 ms. A start bit and three preamble bits have been added to synchronize the reader (where necessary) and to identify the type of data, respectively. This results in a total code length of sixteen bits.

Figure 4
Format of the inserted data.



5.2. Conveying the code in analogue video

For analogue video applications, there are already several existing standards which can accommodate data; for example, teletext, International Insertion Data, Insertion Test Signal and Vertical Interval Time Code formats. These have all been considered for this application but, unfortunately, are unsatisfactory for various reasons.

In the analogue domain, the simplest method is to place the data on a spare line within the vertical blanking interval (VBI) of the video signal. Space in the VBI is at a premium after teletext has been added for transmission, but is normally available earlier in the broadcast chain, where this delay code system will be used.

For analogue video a simple format was therefore produced as shown in *Fig. 4*.

The choice of which line to use for the code is not important; it will depend on which lines are available in a particular installation.

The data carried by this system does not require heavy protection against errors, as the response time of the compensating audio delay will tend to filter out the occasional incorrect code. A simple confidence-counting process on continuously-transmitted codes offers sufficient error protection.

The bit rate chosen is 0.844 Mbit/s as this simplifies the implementation of the system in a digital environment; it results in a bit width which corresponds closely to 16 component samples (13.5 MHz sampling) and 21 composite samples (17.73447 MHz sampling).

An off-screen photograph showing the delay code within the VBI is shown in *Fig. 5*.

5.3. Using the code with digital video

Although the main emphasis of this description has been directed towards analogue installations, the technique is equally valid for digital environments, the only difference being the method employed to convey the delay data.

The digital video standard, ITU-R Recommendation BT.656 [2], includes provision for ancillary data and this facility could be used directly to carry the data values within the digital video signal. However, this may not be the most appropriate method, because other data areas are available, or a direct representation of the signal shown in *Fig. 4* could be used. This latter technique, although wasteful of bit capacity, avoids the need for special transcoding between formats and is currently being proposed for timecode. It is therefore not clear, at present, which will be the best method to convey the data in a digital environment, or the format to be used [3]. The technique described here is, however, equally valid and otherwise operates in an identical manner.

5.4. Recording the delay codes

There is no reason why the delay codes cannot be recorded, although VTRs may have to be set to replay the data line in the VBI. However, it would be normal practice to record only corrected audio, so that an isolated machine could be used to play back with correct synchronization.

6. Transparency to the delay code

Equipment which delays the video or audio does not need to pass the delay code, because the code will be read and reinserted in the updating process. However other equipment, such as a vision mixer, must not strip the delay code from the vision signal or render it unreadable. In tests to date, this has not been a problem; the studio vision mixers have all passed the codes in the VBI with little or no reconfiguration being necessary.

During a fade or wipe, many mixers will pass the VBI of the original source until the transition is completed, and then select the VBI of the new source, passing the new delay code to the output as required. It is possible, however, that some mixers will perform a fade through the whole field,

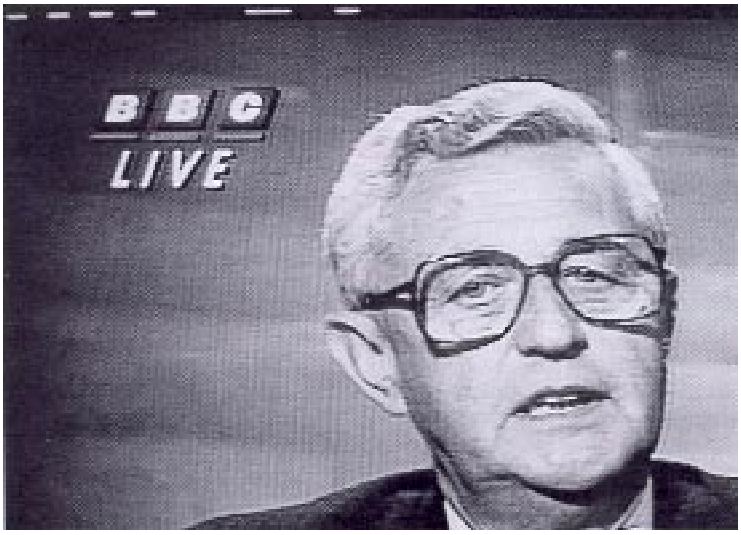


Figure 5
Off-screen photograph from an underscanned monitor showing the delay code in the vertical blanking interval.

including the VBI. If a partial fade or wipe is held, a corrupted data value could be repeated enough times to be interpreted as a valid delay code. A held fade can however be detected easily, as the resulting mixed data levels will encroach into the disallowed region between logic “1” and “0”.

If a mixer *wipes* through the VBI, the situation is likely to be detected by looking for unexpected logical transitions and any erroneously-generated values would have to be repeated several times to pass the confidence counting check and be accepted. This is only likely to happen if a wipe is held stationary, but could be detected by sending the code twice on the same line, whereupon a held wipe would almost inevitably result in two different data patterns.

Initial enquiries have also been made to establish the transparency of digital video products to ancillary data. As product development is generally still at an early stage, many manufacturers have not considered how best to handle such additional data. It is nevertheless expected that this issue will be addressed as developments continue, and transparency to ancillary data will be achieved in most products as applications demand.

7. Results and future installations

To test the feasibility of these ideas, six prototype delay code inserters were built and included a variety of interfaces to allow connection to video and audio processing equipment, and audio delays. After initial tests in various studios, these were installed in the BBC’s main national news studio at Television Centre, London. Five units provided correction for four synchronizers and a DVE. The sixth unit was used to drive an audio delay device which had been optimized to allow rapid and inaudible delay changes.

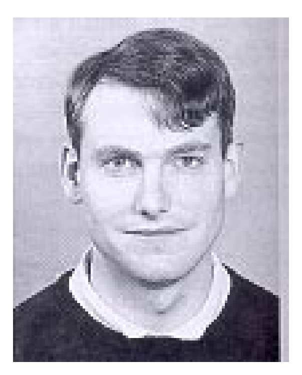
The system worked as expected. The delay codes inserted at the output of each synchronizer and the DVE were routed through to the audio delay device which corrected the delays, rapidly and unobtrusively, at the appropriate times. The corrected audio was used on air, and provided a worthwhile improvement to the quality of the programme, even for such a limited installation.

Two unanticipated operational difficulties came to light during the field trial, both of which were caused by monitoring feeds derived downstream of the compensating audio delay. In the first case, the audio monitoring feed in the studio control area was taken from the output of the audio delay device and was picked up by the talkback microphones and hence heard by the newsreader in his earpiece. When an effect was introduced, or an interview carried out using a remote source, the system tracked the extra delay and hence delayed the audio heard in the control area. The newsreader then heard himself, now delayed, in his earpiece. This is unacceptable for anything other than short compensating delays.

The second instance where this problem occurred involved a remote studio which fed into the news studio. The output of the news studio was monitored at the remote studio, for confirmation that its signal was in use and being received correctly. The necessary reverse audio feed, taken from the output of the news studio, had been delayed to maintain correct synchronization with the video from the remote studio, which had passed through a synchronizer at

__David Kirby__ joined the BBC’s Research Department after graduating from Cambridge University, England, in 1977 with a B.A. degree in Electrical Sciences. After working briefly on quadraphony and techniques for improving the quality of television sound, he became involved in video related projects and, in particular, the development of a disc-based real-time video storage system for animation work. Following this he returned to the field of audio and the application of disc-based storage for editing.

More recently, Mr Kirby has been involved with a variety of other digital audio projects and the development of techniques for maintaining the synchronization of sound and video in television production. He is also working on multichannel sound for enhanced television systems and the assessment of the various ISO/MPEG-2 low bit-rate audio coding proposals currently under development.



__Matthew Marks__ graduated from University College London (UCL) in 1990, having spent a year as a Pre-university Trainee at BBC Research Department. He has been involved with video signal generation, iso-frequency FM transmission, and digital audio dithering and format conversion.

Mr Marks has also taken part in research into the concepts of the audio/video synchronization project, and has undertaken the development of its hardware and software.

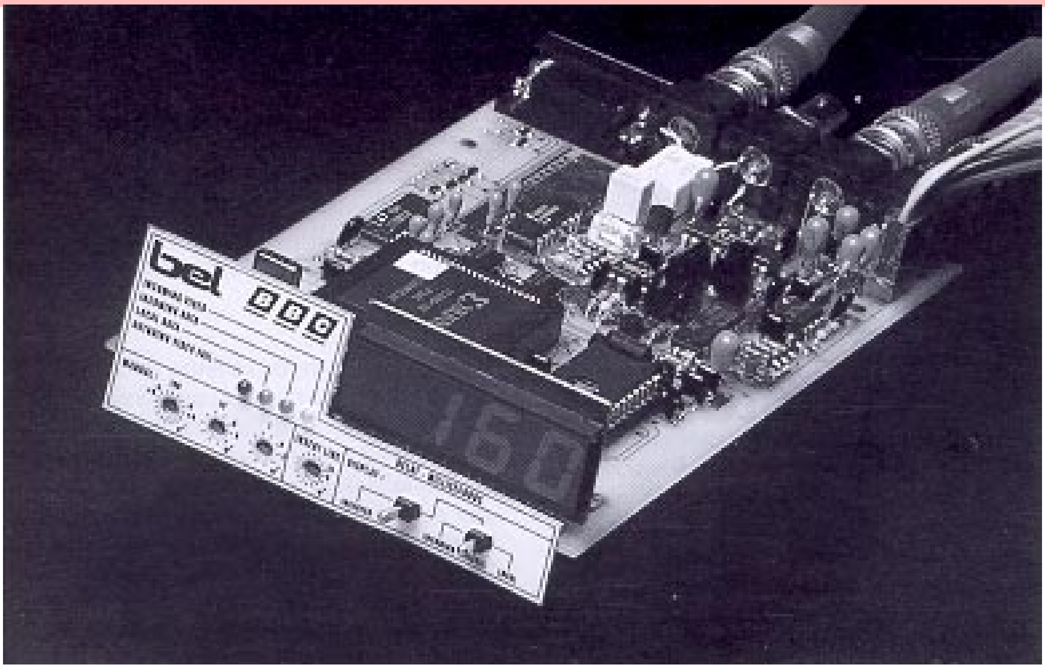


Figure 6
Production version of the delay code reader/inserter.

the input to the news studio. This situation was disconcerting for the remote source, although it did not represent the true situation; the sound and picture were correctly timed at the news studio output.

It should be noted that these problems can occur with other synchronization strategies; they are not inherent to the system described here. Any potential difficulties can be avoided by positioning the compensating audio delay away from the studio and closer to the network transmission point, beyond the point from which all reverse feeds are taken. It is this configuration which will be used in future.

Following this field trial, a wider installation of the equipment is now under way at the BBC's studios in Television Centre. A reduced-cost version has been developed and 35 of the new units, shown in *Fig. 6*, have been manufactured for installation in the news studio, the central equipment area which handles all incoming circuits, the network transmission area and up to three other studios. It is hoped that this widespread installation will cover most instances where serious timing errors are encountered. In addition, a set of equipment for outside broadcast use will allow more complex OB events to be corrected in the same way.

8. Benefits of standardization

The proposal outlined in this paper has also been put forward, through the EBU and ITU-R, 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. It can be seen from *Fig. 6* that the circuitry required for these functions is very simple and could easily be incorporated, by manufacturers, into their own equipment designs. This would then avoid the need for any external units and, as many of the elements may already be present in the equipment, this could be at little extra cost.

Bibliography

[1] CCIR Recommendation 717: **Tolerances for transmission time differences between the vision and sound components of a television signal.**

[2] ITU-R Recommendation BT.656: **Interfaces for digital component video signals in 525-line and 625-line television systems operating at the 4:2:2 level of Recommendation 601**.

[3] Forthcoming recommendation of ITU Task Group 11/2: **Format of ancillary data signals carried in digital component studio interfaces**.