# Managing Audio Delays and Lip-Sync for HDTV

**Rowan De Pomerai**
*BBC Future Media & Technology*

**Viewers who have invested in high-quality HD receiving equipment have problems with audio-video synchronization in the home – due to the processing delays introduced by modern LCD and plasma displays. Set-top boxes often provide controls to manage this, but viewers have no way of calibrating the settings easily.**

**Consequently, the BBC set out to help them by broadcasting a synchronization test signal, as well as undertaking a full review of multi-channel audio in the production and broadcast signal chains.**

High-Definition (HD) television is becoming increasingly popular around the world and, to augment its higher resolution pictures, it often comes with multichannel audio or surround sound. The BBC launched a trial HD channel in May 2006, which became a full service in December 2007. BBC HD, in common with all current HD services in the UK (Sky and Freesat on satellite, and Virgin Media on cable), uses ***Dolby Digital*** as the primary emission audio codec – both for stereo programmes and, where available, 5.1 surround sound.

***Dolby E*** is used by the BBC for delivering audio through the production and distribution signal chains – for transporting the larger number of audio channels and metadata which are required for Dolby Digital transmission. Some months after the start of the broadcasts, however, the technology was still proving temperamental in use; while, generally, we were broadcasting without major problems, small troubles such as audible artefacts at junctions and audio-video synchronization issues were cropping up and, occasionally, larger problems caused more high-profile on-air difficulties. A review of the use of Dolby E throughout our workflows was therefore initiated, with the aim of examining the problem areas and resolving the issues we were encountering.

Additionally, viewers who have invested in high-quality HD receiving equipment have problems with audio-video synchronization *in the home* – due to the processing delays introduced by modern LCD and plasma displays. Set-top boxes often provide controls to manage this, but viewers have no way of calibrating the settings easily, and so the BBC set out to help them by broadcasting a synchronization test signal.

This article looks at the difficulties surrounding audio-video synchronization (lip-sync) through the delivery chain and in the home. Further information on these issues and a full discussion of the other outcomes of the Dolby E review – such as those involved with metadata, system timing and monitoring – will shortly be published in a BBC Research & Development white paper at:

**http://www.bbc.co.uk/rd**.

# Broadcast infrastructure in the BBC

Much of the difficulty in managing the implementation of a new technology across wide areas of the broadcast infrastructure lies in co-ordinating the efforts of disparate technical departments. In the modern BBC, as with many other broadcasters, the problem is compounded by the fact that large swathes of the distribution infrastructure are handled by external organizations. The path of a live high-definition outside broadcast (OB) is summarized in *Fig. 1*, while the corresponding companies involved are illustrated in *Fig. 2*.

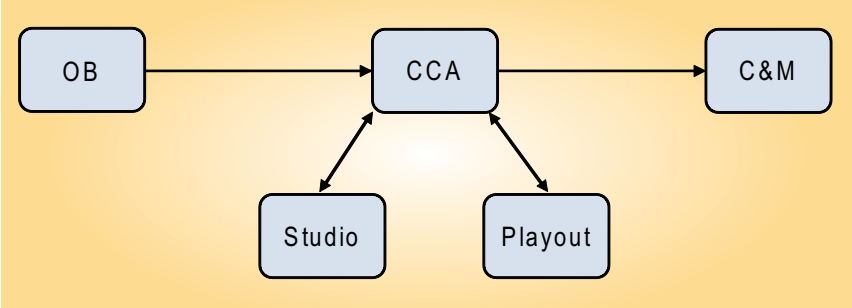


**Figure 1**
**An overview of the signal path followed by a BBC HD OB**

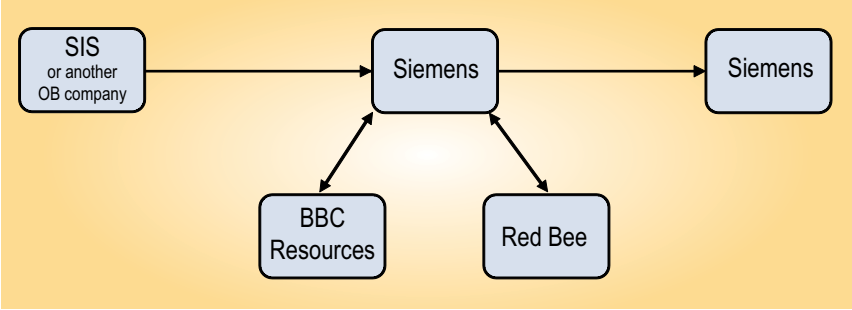


**Figure 2**
**The companies who run the various parts of the signal path**

The major areas can be summarized as follows:

- **Outside Broadcast** – run by one of many OB companies, one common example being SIS Live, formerly BBC Outside Broadcasts. Additionally, a separate links provider may be involved.
- **CCA** – the Central Communications Area is in BBC Television Centre (TVC) and acts as the central routing point of the distribution.
- **Studio** – in the case of an OB insert into a studio programme, a studio in BBC Television Centre may be involved.
- **Playout** – Red Bee Media runs the BBC's playout area, where schedules are managed, junctions are mixed, continuity announcements are added and pre-recorded programmes are played out.
- **Coding & Multiplex (C&M)** – the coding and multiplex operation for digital TV is run by Siemens and located in TVC.

Only through working with all these partners can a coherent strategy be formed for the correct handling of Dolby E to avoid problems.

# An introduction to Dolby E

Dolby E is a data-stream, designed by Dolby Laboratories, which carries up to eight channels of digital audio within a standard stereo channel (AES3), as well as transporting metadata which describes the audio and its reproduction. It is a professional data-stream, designed for use within production and broadcast infrastructures – but not for use as an emission codec or by consumers. It employs light data-rate reduction, ensuring that multiple encode-decode cycles are possible.

Dolby E is often used in conjunction with Dolby Digital, which is a consumer data-stream, also carrying multiple channels of audio (up to six) and associated metadata, and is used as the emission codec by many broadcasters.

The *consumer* metadata used in a Dolby Digital stream is carried by the Dolby E stream, along with some additional *professional* metadata. Therefore the consumer metadata can be transferred from a Dolby E decoder to a Dolby Digital encoder, allowing metadata continuity from studio to home.

Dolby E divides audio into frames at a rate aligned with the associated video. A Dolby E frame may not be split or modified in any way – such as gain adjustment, sample-rate conversion or equalisation. An incomplete or corrupt frame will not reproduce the intended audio, but will likely cause a mute for the duration of the frame or be interpreted as PCM audio, causing a loud audible "splat" in the output. Each frame *(Fig. 3)* is slightly shorter than a video frame, allowing a guard band to be used, and leaving some unused time at the start and end of the frame. The switching point used by mixers and routers falls in this guard band, allowing switching between video sources without the corruption of the embedded Dolby E.

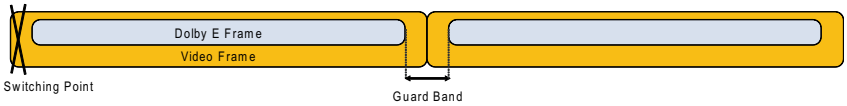


**Figure 3**
**Simplified Dolby E timing (not to scale)**

Significant problems can be encountered where equipment alters Dolby E frames, or where Dolby E is misaligned with the video, due to differing processing and propagation delays associated with audio and video. Seemingly innocuous devices, such as frame synchronizers, can cause problems if not managed correctly and, while Dolby E's frame-synchronous nature allows cut transitions without decoding, actions such as mixing audio and adding voiceovers requires careful implementation of a decode-process-encode cycle. The details of such issues and recommendations of best practice are outside the scope of this article but will be discussed in the associated white paper.

# Synchronization in the delivery chain

Dolby E encode and decode cycles each incur a fixed delay of 1 video frame (e.g. 40 ms in a 25 Hz system). This means that delays have to be carefully managed, as significant audio-video synchronization problems would otherwise occur, although the "round number" nature of the processing delay makes the effects easier to manage.

In a PCM environment, it is expected that programmes will be delivered to a broadcaster with audio in sync with the video. However in a Dolby E system, the choice is less obvious. There are two main options:

- **In-Sync Encoded** – The encoded audio lies on tape (or appears on the link) in sync with the video. The decode delay must be compensated for at the decode site by the use of an equivalent video delay.
- **Advanced** (decode-compensated) – The encoded audio appears one frame ahead of the video, meaning that after the decode delay, the audio is in sync.

The BBC began by using the latter option but, in conjunction with other UK broadcasters, chose to change to the former. This change was motivated by a variety factors, including the ability to make

cut edits without an audio offset. However, the review of Dolby E found that an administrative error had meant that one technical area was still working to the older standard, causing BBC HD to have one frame of constant lip-sync error. This was quickly corrected but is a cautionary tale to those managing Dolby E; delays must be carefully considered!

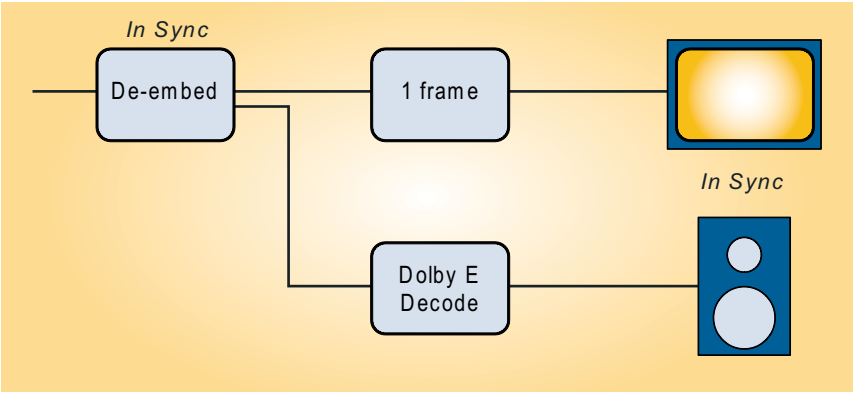


**Figure 4**
**The BBC's chosen delivery and delay solution for Dolby E**

Another potential pitfall is ensuring no double-compensation. Many devices which are Dolby-E-aware, "helpfully" include a frame of video delay to compensate for the audio decoding delay. Such devices include de-embedders, MPEG video encoders and more. Clearly one must ensure that all such delays are known about. Generally such delays will be configurable, but the settings should be checked, because if a de-embedder has a frame of delay turned on and then a discrete frame delay is also used, the video will be double-delayed and hence out of sync with the audio. The BBC often uses video delays in embedders and other equipment to compensate for audio encode/decode delays, but crucially the corresponding audio and video delays are always co-sited, so that a signal leaving any individual technical area (a studio, CCA, playout etc.) are in sync.

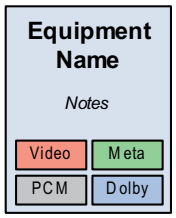


**Figure 5**
**A generic schematic block identifying the delays**

It is important to note that most broadcasters' delivery chains accept mixed content. The BBC's playout area must accept programmes incoming with 5.1 Dolby E audio and those with stereo PCM, outputting all programmes encoded as Dolby E (5.1 or 2.0) with appropriate metadata and correct sync. Therefore the different delays through devices must be considered for Dolby E, PCM audio, video and sometimes separate metadata. In attempting to track these delays, it quickly became apparent that traditional schematics would not present the required information concisely enough. Hence, a simple but effective notation was developed whereby any device can be represented as a generic block with delays indicated for the previously mentioned four elements *(Fig. 5)*.

*Fig. 6* shows this block in use to describe a section of the BBC HD playout systems. This area is an excellent example of the challenges involved with managing sync. The Dolby E decoder, for

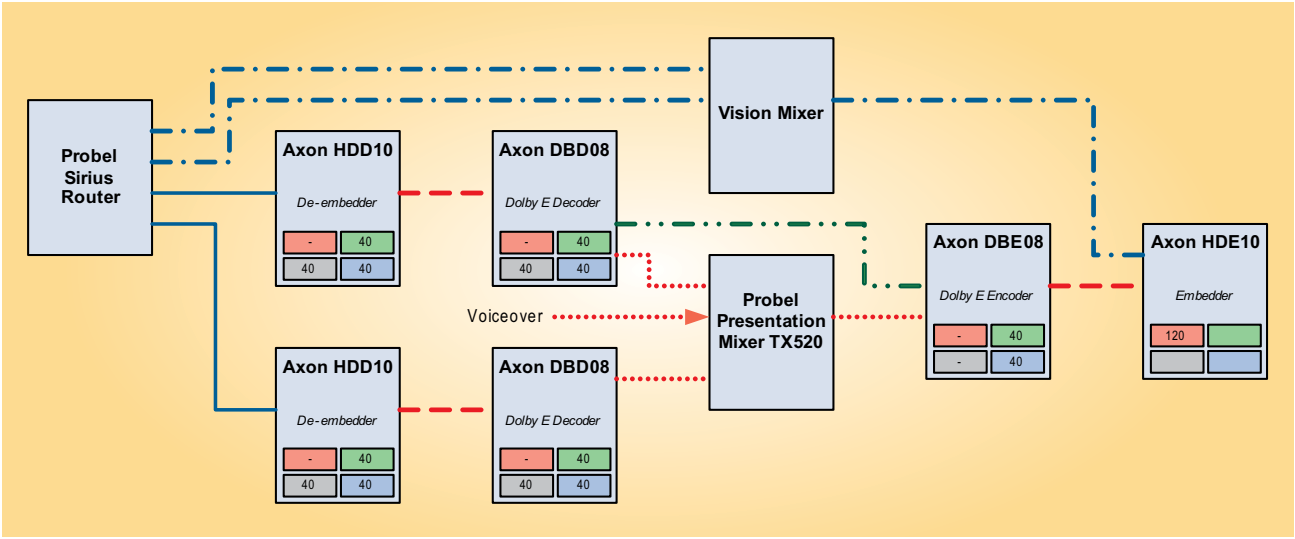


**Figure 6**
**A section of the BBC's playout infrastructure**

example, must be carefully configured such that it passes through the PCM audio unaltered, but delays it by the same amount as the device would delay a Dolby E stream, i.e. 40 ms. If the decoder is incorrectly configured, it will leave PCM audio out of sync with the video if the rest of the system is designed to keep Dolby E in sync.

Meanwhile, the metadata for a Dolby E programme is taken from the decoder to the encoder on a serial link, so any audio delays that occur in between these devices – i.e. in the mixer – must be considered as they must not put the audio out of sync with the metadata. This could cause the metadata to be re-associated with the wrong frame of audio, causing unpredictable effects at junctions. (The mixer's delay is low enough to be negligible in this particular case.)

Finally, the video must be delayed to match the delays in the audio chain. This delay is applied in the embedder, such that the embedded output is in sync. Whilst these issues may initially seem complex, the notation used allows simple arithmetic checks to ensure that the delays for PCM, Dolby E, video and metadata all match.

There is no substitute for testing, however. A test such as Probel's VALID/VALID8, a tone & flash test (from various devices including Tektronix monitoring equipment) or other appropriate sync tests, should be used to ensure correct sync. When doing so, if the test signal needs to be encoded to Dolby E after generation and/or decoded to PCM for measurement, then the delays involved in this must be carefully accounted for, so that the readings obtained are fully understood. The diagrams produced of the BBC's infrastructure, and the delays involved, were used to check and verify that the test results were as expected, never as a substitute for testing.

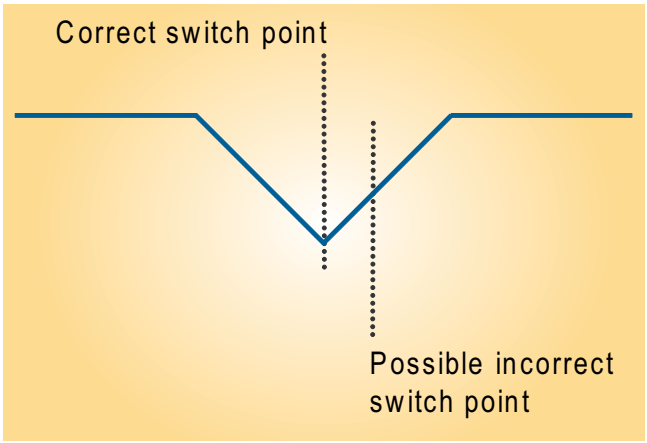


**Figure 7**
**Switching point problems in V-fades**

An additional problem caused by the delays of Dolby E equipment, which may not be immediately obvious, is the latency between the router and the mixer. *Fig. 7* shows two devices each introducing a frame of delay between the router and the mixer. Because the automation systems work in conjunction with the router and mixer to create the transitions used at programme junctions, the effects of this delay can be problematic.

In a V-fade transition, the mixer reduces the volume of the source on the programme bus to zero (silence), then the new source is switched onto the bus and the mixer increases the volume to full level. The intended effect is that the outgoing programme fades out to silence and then the incoming programme fades in. However, the switch from the outgoing to incoming sources happens in the router at the same time as the mixer reaches the silence point. If the mixer immediately begins to fade up again, the outgoing source will be heard once more, as the latency of the de-embedder and decoder means that the effects of the router switch will not propagate to the mixer until 80 ms later. The viewer hears a brief segment of the outgoing audio after they should, before the new programme's audio cuts in, part-way through the fade up. Seen from the point of the mixer, the switch appears to have happened 2 frames too late.

The solution to this problem is the *switchback delay* parameter of the mixer, which effectively turns V-fades into short U-fades. In other words the silence is held for a fixed amount of time, waiting for the router source switch to propagate to the mixer before increasing the volume. This is an acceptable compromise, and is the option now used by Red Bee Media on behalf of the BBC.

Other problems of this nature are discussed in the forthcoming BBC Research & Development white paper on the use of Dolby E but, hopefully, it has been illustrated here that the challenges inherent in adding processing delays to a signal chain may be more complex than initially anticipated.

## Synchronization in the home

Whilst mixed audio formats in the delivery chain create lip-sync challenges for the broadcaster, the viewer at home has a challenge of his or her own. In a situation where a set-top box (STB) is separately connected to a display and an audio receiver or amplifier (by HDMI and optical, for example), the audio and video signal chains become separated. Processing delays in one chain can therefore cause synchronization problems and, with modern LCD and plasma displays routinely adding anywhere up to 100 ms of delay due to deinterlacing and other processing, the user can be left with significant synchronization errors.

Worse still, the video is late with respect to the audio. With some simple consideration of the properties of sound and light waves, it can be seen that this is more disconcerting for the viewer than audio being late. Sound travels slower than light, so humans are used to the sound of an event reaching them fractionally after the visual. Anyone who has been to a concert in a stadium, park or other large venue will know that the sound of the singer and the sight of their lips moving on stage (or on video screens) are noticeably out of sync, with the audio being late. So introducing a situation in the home where the reverse is true – the video lags behind the audio – creates a highly unnatural effect which is confusing to the brain.

Work by BBC Research & Development staff has investigated the difference in perception of audio-video synchronization between standard definition and high definition. Early results seem to indicate that some subjects are more sensitive to lip-sync in HD, highlighting the need to minimize sync problems. The work is presented in AES Convention Paper 7518, *Factors affecting perception of audio-video synchronization in television* by Andrew Mason and Richard Salmon.

Many STBs have a configurable delay on their optical audio outputs, allowing the audio to be delayed by increments of 20 ms (half a frame). However with no reference signal upon which to base their judgements, viewers are left relying on lip-synchronization in television programmes, which is subjective, inaccurate and of course could even be wrong at the point of broadcast. The BBC has strict guidelines as to acceptable lip-sync but, by necessity, they are finite. Anything within 10 ms audio early to 20 ms audio late is allowable; this is unnoticeable to the viewer in normal circumstances, but insufficient for sensitive alignment of equipment.

## BBC sync test

Given that BBC HD only broadcasts programming at certain times of the day, a decision was taken that a sync test could be broadcast as part of the promotional loop which plays at other times. Once every two hours during the daytime, 90 seconds of a sync reference is played in order to allow viewers to adjust their equipment. The design of this test, and the process of ensuring that it was broadcast in sync to much tighter tolerances than usual, was the subject of considerable effort.

There were some unique requirements for a sync test to be broadcast in this manner. First and foremost, it should be as easy as possible to perform the test “by eye” without specialist equipment. Additionally, in order for ourselves to be able to perform precise measurements, automated millisecond-accurate measurements were a requirement. Finally, it would be ideal if the measuring device required to perform these precise measurements was electrically simple. This would mean that testing devices could be produced in-house, and potentially the details for how to build such devices could even be made available to the public.

The starting point was work from BBC Research & Development, who had previously developed a sync test sequence referred to as the *digital clapperboard*. *Fig. 8* shows a still frame from the modified BBC HD version of this sequence.

The signal consists of three primary elements. The top right of the image is a bar which extends down from the top of the screen until it touches a static line just above the centre of the image. The extending bar touches the line at the same time as when an audio “snap” is heard, then retreats back to the top of the display. The second element consists of three white horizontal lines, five video

lines each, which appear for one frame at the same time as the audio snap. Finally there is a horizontally expanding bar moving across a time scale marked in video frames. When correctly synchronized, it reaches the centre of the screen at the time of the audio snap.

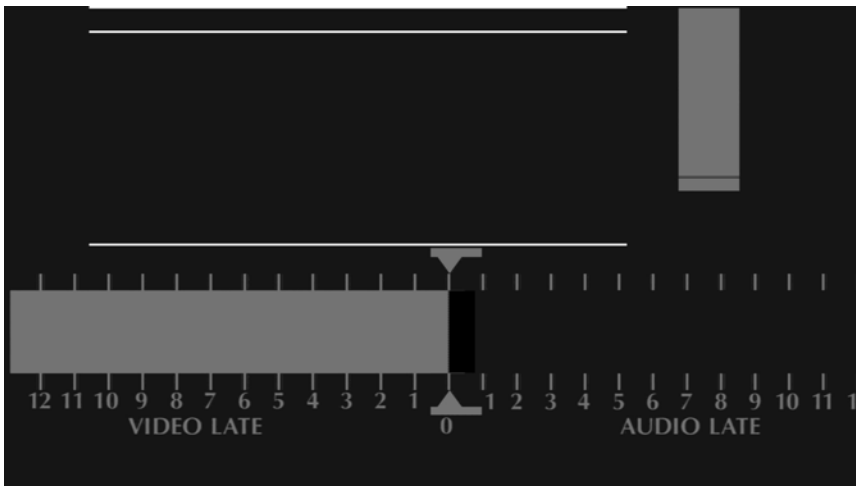


**Figure 8**
**The BBC HD sync test**

The horizontal scale is best used for measuring by eye. One can examine the right-hand tick mark (11 frames audio late) and note whether the snap has been heard before the bar reaches this point. Assuming it has, move on to the next tick mark to the left, and perform the same test. Continuing in this manner, eventually a point will be reached where the visual bar touches the tick mark at approximately the same time as the audio snap, or slightly before it. Most users can identify the sync offset to a precision of half a frame (20ms) precision using this method.

It is worth mentioning at this point, the origin of the audio snap. The sound used is a recording of two blocks of wood being banged together. Following a series of tests at BBC Research & Development, this was found to be the sound which was most effective for subjective sync tests, largely due to the strong transients at the start. Had the primary objective of this test been automated measurement, it would perhaps have been easier to use an artificially-generated tone for example; however, the requirement for "by eye" measurement made this sound the best choice.

Clearly, the broadcast of a sync test on one of the BBC's flagship channels requires careful planning to ensure that the test signal reaches the viewer in sync. It would have been an embarrassing error to broadcast a sync test which was itself out of sync. As such, careful testing of the entire delivery chain was undertaken to ensure millisecond-accurate timing of this signal throughout.

An electronic test device was used to provide precise measurements of the sync signal. Using an analogue audio input to measure the audio snap and a "light pen" to measure the horizontal white lines which flash on a display, the offset in milliseconds between the two is derived and displayed. Because this measuring method uses a display and a decoded version of the audio, the delays of these devices must be carefully understood. A CRT display must be used due to the delays involved even in professional LCD and plasma displays, and the frame of delay from the Dolby E decoder (if the signal is Dolby E encoded) must be accounted for.

The first challenge was to get the test signal itself inserted into the promotional loop with perfect synchronization. The sync test signal was produced and laid to tape. It was then ingested in the edit suite, and inserted into the promotion by the editor. Once the video edit was completed, the promotion was transferred to the dubbing suite for the audio finishing. From there the whole loop was Dolby E encoded and laid back on to tape (with a frame of video delay used to compensate for the Dolby E encode, of course). This tape was taken back to the edit suite where it was played

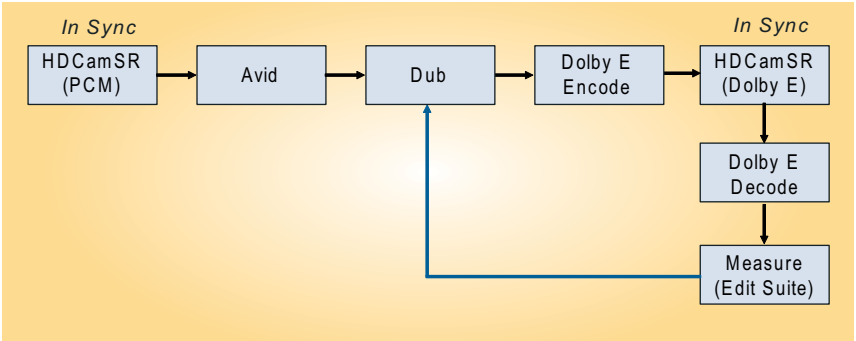


**Figure 9**
**Producing a sync version of the HD promotional loop**

and measured, producing a sync error value which was taken back to the dub in order for the audio offset to be adjusted accordingly. The reason for this back-and-forth action was that an HD CRT was only available in the edit suite (so reliable measurements could only be made there), whereas the only place where sub-frame timing adjustments could be made was the dubbing suite, whose systems allow the audio to be moved in samples. Eventually, a full copy of the promotional loop was available on-tape and measured to be in sync.

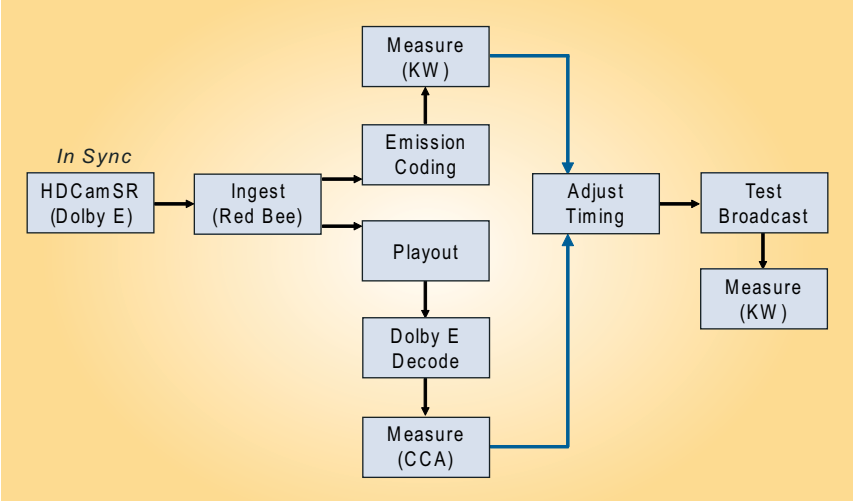


**Figure 10**
**Measuring sync in the broadcast chain**

Next, the playout chain was tested. The tape was ingested into the playout server, and various measurements taken along the playout signal chain, within the playout centre and down-stream in the Central Communications Area. The emission coders could not be tested on-air as this would potentially involve broadcasting the signal out of sync, so a duplicate set of encoders at BBC Research & Development's Kingswood Warren (KW) site were used to perform the measurements. The playout and delivery chain was found to be in sync to a tolerance of less than 2 ms, although the emission coders did introduce some offset. The measured value of this offset was used to adjust the video delay applied in the on-air coder, and we were ready to broadcast the test.

There is an old adage in broadcasting that people claim a signal was "alright leaving me", implying that any error was caused at the receiving end. The desire for the BBC HD sync test was to go a step further and ensure it was "alright arriving at you". In other words, the signal was required to be correctly timed when received off-air. To verify this, the sync test signal was broadcast and measured off-air at Kingswood Warren. This test used the most reliable and precise measurement possible; a transport stream analyser was used to examine the Presentation Time Stamp (PTS) values of the MPEG transport stream, allowing identification of the exact timing of the video flash. Because a PTS covers only one video frame but multiple audio samples, the PTS of the start of the audio snap was not sufficient to make the corresponding audio measurement. However, by examining the decoded waveform, the exact audio sample number within the PTS block could be found and, (knowing the audio sampling rate used), the time difference between the snap and the flash could be precisely measured. The off-air test revealed this offset to be 0.9ms, an excellent result.

BBC HD continues to broadcast this sync test multiple times daily, so audiences now have a way to test the synchronization of their home equipment. The test has been explained in blogs by both myself and Andy Quested, head of technology for BBC HD, and the response from viewers has been overwhelmingly positive. Additionally, it has generally led to a reduction in the number of comments received about the sync of the channel and, while problems on individual programmes do of course occur occasionally, we are confident that the channel's broadcast chain is in sync, and we continue to push our technology partners to verify synchronization any time any change is made to devices in that chain, so that the synchronization remains correct in the future.

## Summary

In the exciting world of High-Definition Television, many of the technologies which have made HDTV possible – such as Dolby E for transporting audio for the broadcaster, and large-screen LCD and plasma displays for the home – cause complexities of timing which are new to both the broadcaster

**Rowan de Pomerai** graduated from the University of York (England) in 2007 with an M.Eng. in Electronic Engineering with Media Technology. He joined the BBC in Research & Development where he worked on computer vision technologies in the Production Magic team, developing innovative interaction systems for television presenters to utilise computer graphics to greater effect.

Mr de Pomerai then contributed to the tapeless production system, Ingex, before taking up the reins of BBC HD's multichannel audio review. As well as gaining a better understanding of the technology involved for the benefit of BBC HD, discussions with the EBU's EHDF subcommittee and other UK broadcasters has led to increased consensus on how different broadcasters can work in compatible ways.

In advance of the BBC's move north to Salford, in Greater Manchester, Rowan de Pomerai will shortly be involved in developing an additional Research and Development laboratory in Manchester, which will work in conjunction with the existing lab to continue leading the next generation of broadcast technology.

and the viewer. Only through careful consideration of the issues involved in the use of such technologies, along with thorough testing of all systems, can a successful implementation be achieved. The results however are worth the time and effort. BBC HD provides top quality content to viewers in more detail than ever before and, with high-quality surround sound, the viewing experience is more exciting and immersive than anything previously available. It has been a challenging journey to make everything work well, and no doubt will continue to be so, but as HD becomes the norm in future, the lessons learned now will be useful for years to come.

## Acknowledgements

The author would like to thank Andy Quested, Head of Technology for BBC HD, for his sponsoring of this project and support throughout. Staff of Red Bee Media, Siemens, BBC Resources and various BBC departments have provided considerable effort and support to the work described here, and suppliers such as Dolby and Tektronix have given their time and advice generously. Finally, colleagues at BBC Research & Development, including Andrew Mason and Trevor Ware, have provided extensive expertise and assistance, as well as the development of the original *Digital Clapperboard* which was the basis of BBC HD's sync test.