



Lip Sync Revisited – Solutions Are in Sight

Previous editions of *TV TechCheck* have reported on the problem of audio-video synchronization, commonly known as “lip sync,” where noticeable errors are seen on television channels distributed on terrestrial broadcasting, cable and satellite. There are various reasons for this, including the introduction of complex digital processing for audio and video that can introduce differential delays for audio and video in production, post and distribution systems.

Standards organizations, including ITU, ATSC, SMPTE, EBU, IEC, AES, and SCTE, have been studying the problem for many years, and there are several standards and recommended practices for tolerances that should be adopted for different parts of the broadcast chain. The [ITU-R BT 1359](#) recommendation states that the threshold of detectability of lip sync errors is about +45 ms to –125 ms (audio early to audio late) and that the threshold of acceptability is about +90 ms to –185 ms, on the average. However, because of the uncertainty of synchronization that may exist in the source material, the ITU recommends that the timing difference in the path from the output of the final program source selection element to the input to the transmitter for emission should be kept within the values +22.5 ms and –30 ms. These distribution tolerance figures are comparable to the recommendations of the [ATSC IS/191](#) finding, which gives a range of +15 ms and –45 ms at the inputs to the DTV encoding devices, while the EBU recommendation [R37-2007](#) gives a range of +40 and –60 ms at the output intended to feed broadcasting transmitters.

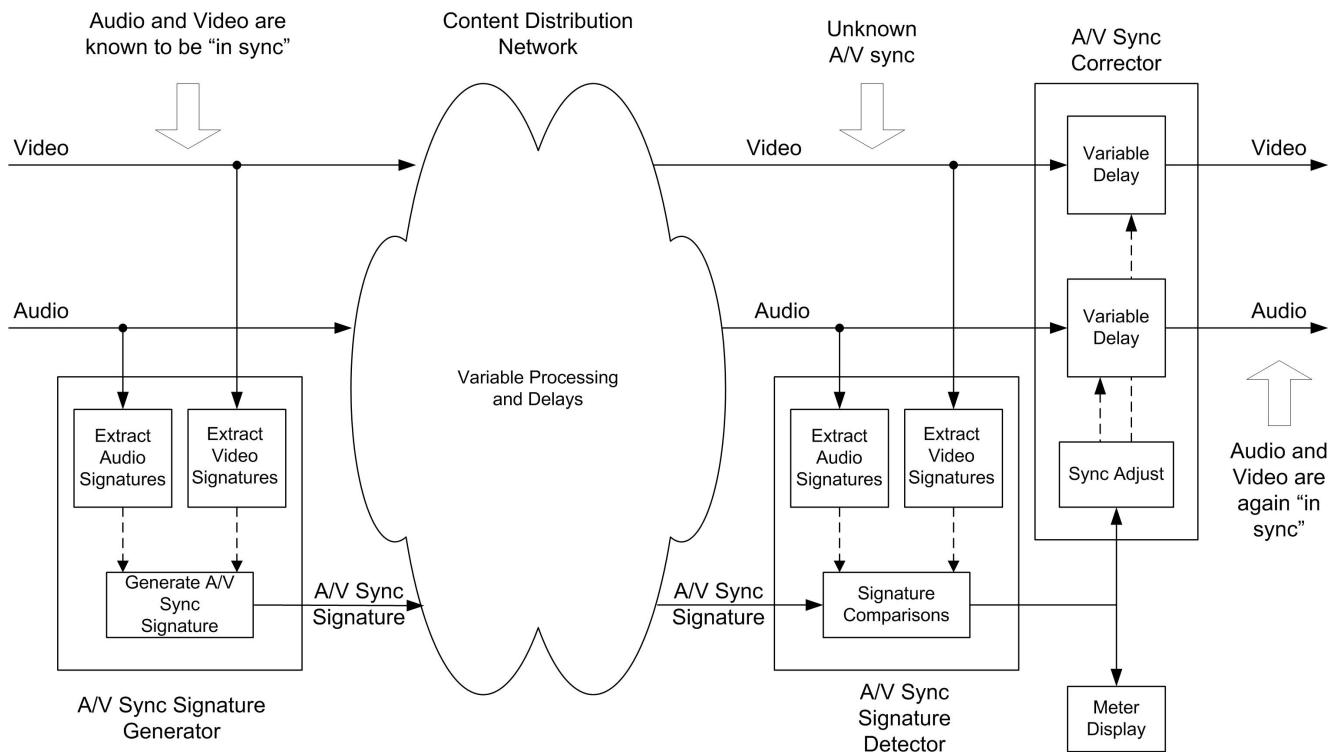
In the above-referenced documents, the ITU states that if correction of errors is not possible, then each downstream segment that is not under the control of the broadcaster shall not introduce any timing error in excess of +/-2 ms, while the ATSC says that designers should strive for zero differential offset throughout the system, and the EBU recommends that the accuracy of synchronization at each stage should lie within the range of audio 5 ms early to 15 ms late. Some broadcasters, and others, have questioned the particular tolerance recommendations mentioned, but there is general agreement on the need to reduce errors to a small value at each step in the broadcast chain.

It has been known for some time that some consumer DTV receiver products can introduce lip sync errors due to incorrect decoding and handling of MPEG signals. As reported in *TV TechCheck* of August 3, 2009, the CEA has addressed that problem (although only for new designs) with a recently published *Recommended Practice on A/V Synchronization Implementation in Receivers*. Apart from that publication, however, as yet, there are few recommendations from the standards organizations as to how the specified levels of synchronization should be achieved through the broadcast chain.

Progress with Solutions for Lip Sync Errors

A prerequisite for A-V sync correction is first to measure the error. It is comparatively straightforward to assess or measure lip sync errors in out-of-service paths using test signals, either of the “beep-flash” type, with equipment available from several vendors, or test signals such as the [Visualizer™](#) Test Pattern from Sarnoff (first mentioned in *TV TechCheck* of October 29, 2007). It is, however, much more difficult to make such measurements in-service, with regular programming, but this is where monitoring and correction is most important. One current product that allows absolute measurements without affecting the program is the [LipTracker™](#) system from Pixel Instruments, but this only works when lip movements are visible and is perhaps impractical to use for all programming. The Tektronix AVDC-100 product was briefly available to detect and correct changes over parts of the program chain using a watermark signal but it was for SD video only and has been discontinued. There have been other products that permit changes in A-V sync to be detected and corrected, but these generally required both input and output versions of the program to be available at one point and were not generally applicable in a distributed system.

At the NAB Show 2009 earlier this year, [Evertz](#) introduced the *IntelliTrak™* system, for in-service measurement and correction of changes in A-V synchronization that may occur within a facility or over multiple links in a broadcast chain. At IBC in September, [Miranda](#) introduced their [Densite HLP-1801](#) product with similar capabilities. Both these systems are based on the concept of analyzing particular characteristics of the audio and video incoming program signals and generating an audio-video sync *signature* or *fingerprint* that can be transmitted as metadata to a downstream point in the program chain. At the downstream point, the audio and video signals are again analyzed and a new signature generated that can be compared with the signature from upstream. This comparison results in a measurement of the change in A-V delay, or lip sync error, between the two points, which can be used to delay the audio or video by the appropriate amount to correct the error. Dolby Laboratories have also been working on a similar technology and Kent Terry will present a paper [Detection and Correction of Lip-Sync Errors Using Audio and Video Fingerprints](#) at the SMPTE Technical Conference in Hollywood later in October. The figure below, taken from the Dolby paper, shows the basic principles involved and is used with permission. It should be noted that the Dolby system is still at the laboratory stage and is not an available product.



Real-Time A/V Signature System Block Diagram
(with acknowledgements to Dolby Laboratories)

Although there are some differences, it is believed that all these *fingerprint* systems have similar capabilities and the signature or fingerprint comparison will work even when the program material has been passed through multiple links in the chain, including addition of bugs and overlays, distribution as compressed bitstreams and other video and audio processing. The method by which the signature or fingerprint is carried from one point to another is not critical, and can be implemented as metadata embedded with the video signal (e.g., in baseband VANC) or through a separate network connection or via the Internet. The other common feature is that they all measure the change in A/V sync from the input reference point, and do not actually measure the absolute sync accuracy of video and audio. Good system design is therefore still required to ensure a high degree of confidence that incoming signals have minimum a-v sync errors.

The Need for a Standard

As they have been developed independently, the algorithms used to produce the A-V sync signature or fingerprint, and the exact form of the metadata are almost certainly different in the three systems mentioned above and will require the equipment at different points in a broadcast chain to be from the same manufacturer.

In the world of broadcasting, however, the entire chain from production to station output is not usually under the control of the broadcaster and it is highly desirable that equipment from different manufacturers should interoperate when used at different points in the chain. Ideally, the form of the fingerprint signal, and its method of carriage would be standardized, such that the fingerprint generated by equipment from one manufacturer, say at a production studio, post house or network release center, will be able to be read by equipment from another manufacturer downstream, say at a broadcast station, and used as input to the delay measurement that can be used to correct any lip sync error detected, before transmission. The same techniques can be used by broadcasters to detect a-v errors introduced in their signals by downstream systems such as cable or satellite.

The SMPTE 22TV Lip Sync Ad Hoc Group, chaired by Graham Jones of NAB, has just started to investigate the possibility of producing a SMPTE standard for the fingerprint signal and/or the method(s) of metadata carriage. Potentially, this could require the manufacturers of existing products for these systems to cooperate to harmonize the fingerprint signals used in their products so they can interoperate. This is much more likely to happen if broadcasters, who are the customers for such systems make known their desire for a single standardized fingerprint system. For those with long memories, this has happened before, when pressure from customers forced Ampex and Sony to modify their similar, but slightly different, 1-inch VTR products to comply with the SMPTE Type C Standard, which for many years was the most widely used professional videotape format.

Readers of *TV TechCheck* who would like to support an initiative for a single Lip Sync Fingerprint standard are encouraged to contact Graham Jones at gjones@nab.org with their comments.

2010 NAB Show Call for Speakers

Call for Technical Papers – NAB Broadcast Engineering Conference



The 2010 NAB Show will host the 64th Broadcast Engineering Conference. This world-class conference addresses the most recent developments in broadcast technology and focuses on the opportunities and challenges that face broadcast engineering professionals. Each year hundreds of broadcast professionals from around the world attend the conference. They include practicing broadcast engineers and technicians, engineering consultants, contract engineers, broadcast equipment manufacturers, distributors, R&D engineers plus anyone specifically interested in the latest broadcast technologies.

Do you have something to share?

If you feel qualified to speak at the NAB Broadcast Engineering Conference, we invite you to [submit](#) a technical paper proposal. Not all acceptable submissions can be included in the conference, due to the large number of submissions that are received and the limited number of available time slots.

PLAN TO ATTEND!

The IEEE Broadcast Technology Society 59th ANNUAL BROADCAST SYMPOSIUM

October 14 -16, 2009

The Westin Alexandria

Alexandria, VA, USA

www.ieee.org/bts/symposium

2009 ATSC Seminar on Audio Loudness

Wednesday, November 4, 2009

Wiley Rein Conference Center

1776 K St, NW

Washington, DC 20006

<http://www.atsc.org/seminars/loudness09.php>

[Official NAB Privacy Policy](#)

© 2009 National Association of Broadcasters 1771 N Street, NW, Washington DC 20036

Cost for ATSC members is \$50.00 for pre-registrants, \$75.00 on-site.

ADVERTISEMENT

