



White Paper

Audio and Video Synchronization: Defining the Problem and Implementing Solutions

Linear Acoustic Inc.
www.LinearAcoustic.com

Introduction

With the introduction of advanced digital delivery systems for audio and video, there is an increased awareness of the timing relationship between audio and video. Owing to advanced data compression technologies such as Dolby Digital (AC-3) for audio and MPEG-2 for video, sound is clearer and pictures are sharper. Technologies such as Digital Television (DTV), DVD, Direct Broadcast Satellite (DBS), and Digital Cable use these compression techniques to deliver extremely high-quality programming to consumers. However, it is a misalignment of these same systems that is the root cause of most audio/video synchronization problems. Perhaps this misalignment is due to a lack of understanding how the system functions as a whole.

It may be useful to define some benchmarks of proper audio/video synchronization before attempting to identify the problems, their causes, and their solutions. In this manner, we will have a measure to which any point in the signal path can be compared and judged. Hereafter, we will refer to Audio/Video Synchronization as A/V Sync.

Audio/Video Synchronization Measures

Most film editors are able to detect A/V Sync errors as short as +/- ½ film frame. As film is projected at 24fps in the US and 25fps in Europe, this equates to approximately +/- 20msec. It is claimed that some editors can detect even smaller errors, but this might be more accurately attributed to their familiarity with the material being viewed. Other figures abound, such as +/- 1 video frame (+/- 33-40msec). Dolby Laboratories has specified that any Dolby Digital decoder must be with the range of +5msec audio leading video to -15msec audio lagging video. This is because human perception of A/V Sync is weighted more in one direction than the other.

It is a fact that light travels much faster than sound. We are all used to seeing this proven, although as it is such a common situation many times we do not notice. For example, a basketball hitting the court in a large sports venue would appear relatively correct to the first few rows, but the further back a viewer gets, the more the sound lags behind the sight of the ball hitting the floor. The further back you get, the more the sound lags, but it still seems OK.

Now, imagine if the A/V timing was reversed. You are watching a basketball game, and the sound of the ball hitting the court arrives before the ball looks like it makes contact. This would be a very unnatural sight and would seem incorrect even if you were in the first few rows where there was just a small amount of A/V Sync error. The point is that the error is in the “wrong” direction. In summary, human perception is much more forgiving for sound lagging behind sight as this is what we are used to seeing in everyday occurrences.

The International Telecommunications Union (ITU) released ITU-R BT.1359-1 in 1998. It was based on research that showed the reliable detection of A/V Sync errors fell between 45msec audio leading video and 125msec audio lagging behind video. That was just for detection, while the acceptability region, and therefore the recommended maximum was quite a bit wider. In summary, the recommendation states that the tolerance from the point of capture to the viewer and or listener shall be no more than 90msec audio leading video to 185msec audio lagging behind video. This range is probably far too wide for truly acceptable performance, and tighter tolerances are generally obeyed.

Delays in the Television Plant

A/V Sync issues within the TV plant are not new to digital television, they are perhaps more noticeable. Although the obvious may be re-stated, there is enough of a problem that still exists in NTSC that it is worth the trouble to identify and fix them prior to beginning DTV operation. Unlike

Some basic points to keep in mind are that in general, audio operations are very low latency. Compression, equalization, mixing, etc... can typically be accomplished in under 1msec in the digital domain, falling to microseconds in the analog domain. Generally, no compensating video delay needs to be added as the latency is so low.

Video processing, on the other hand, takes substantial amounts of time, usually no less than one video frame. Similar to audio, any time a video signal is digitized, operations upon that signal will take longer. As most video effects are unable to be performed in the analog domain, delay is inevitable.

It is interesting to note that processing audio and video signals has the opposite of the desired effect. As video processing takes longer, the video signals will be delayed with respect to the audio signals and A/V Sync will seem incorrect much sooner. It is critical that compensation is provided for any video device that has a delay in excess of a few milliseconds. An equal amount of delay should therefore be applied to the audio path.

The ITU has made a further recommendation, ITU-R BT.1377 that makes the logical suggestion that audio and video apparatus is labeled to indicate processing delay. This delay should be indicated in milliseconds to avoid any frame rate discrepancies, and if the delay is variable the range should be stated. In the case of variable delay, a signal that can control an audio delay should also be provided. By following these recommendations, it is apparent that regardless of the actual delay, compensations can be made and A/V Sync ensured.

A few typical operations are presented below. Two are common pieces of any video facility and have simple, logical solutions. The third is a somewhat surprising source of sync error that probably does not require any changes, but should be taken into account during facility design and troubleshooting.

Video Frame Synchronizers

Due to its very nature, a video frame synchronizer causes between one and two variable frames of delay. In this case, a special audio delay that is able to track the variable delay of the frame synchronizer is required. Most video frame synchronizers are available with matching and tracking audio delays and should always be purchased as a set as there is currently no standard interface that represents A/V Sync values.

Digital Video Effects

Digital video effects can add from one to many video frames. As the delay of a DVE is generally a fixed value, a fixed value audio delay can also be used. Devices with fixed delays are easier to deal with if they are always kept in line, or if they must be removed then a fixed video delay equal to that of the device is inserted in its place. This will prevent having to dynamically adjust audio delay and create an audible disturbance.

The Camera and Display

During experiments conducted in Australia by the ITU, it was found that Tube cameras take approximately 20msec longer to produce a frame of video than CCD cameras. In retrospect, it is obvious why such a difference would exist. Tube cameras scan an image a line at a time, and do so twice to produce a full frame of video. CCD cameras can capture the entire image at once, and as there is no scanning per se, convert it to electrical signals faster.

This difference is of no consequence if each signal is displayed in a compatible manner. For example, if the output of a tube camera is displayed on a CRT one line at a time, twice for each frame, there is an overall delay, but it is fixed. The same is true for the output of a CCD camera reproduced by a full-field flat panel display. Issues can arise when the output of a CCD camera is displayed on a CRT as the image is captured at once but displayed sequentially.

Although these differences are relatively minor and depend on the consumer's display, they should be taken into account when designing or troubleshooting a television plant. In an all-CCD camera plant, the delay will be of little consequence. In a plant that mixes cameras, it can be a source of errors that will vary depending upon which camera is active. It is doubtful that these differences need compensation, but it is worth at least knowing where errors might creep in.

A/V Sync in the MPEG-2 System

The MPEG system provides the proper tools to make A/V Sync absolutely correct. Each audio and video frame has a Presentation Time Stamp (PTS) that allows the decoder to reconstruct the sound and pictures in sync. These PTS values are assigned by the Multiplexer in the MPEG encoder. The decoder receives the audio and video data ahead of the PTS values and can therefore use these values to properly present audio and video in sync. It is imperative that audio and video are applied to the Dolby Digital (AC-3) and MPEG-2

encoders *In Sync*. A very common mistake is to calibrate the multiplexer to compensate for plant differences. Although this may work fine in the short term, it should be avoided in permanent installations. Larger problems will likely result, including issues with some consumer decoders.

Figure 1 is a simplified block diagram of a typical MPEG-2 video encoder, Dolby Digital (AC-3) audio encoder and a multiplexer. Note that the Dolby Digital (AC-3) encoder can be either internal or external to the video encoder and multiplexer

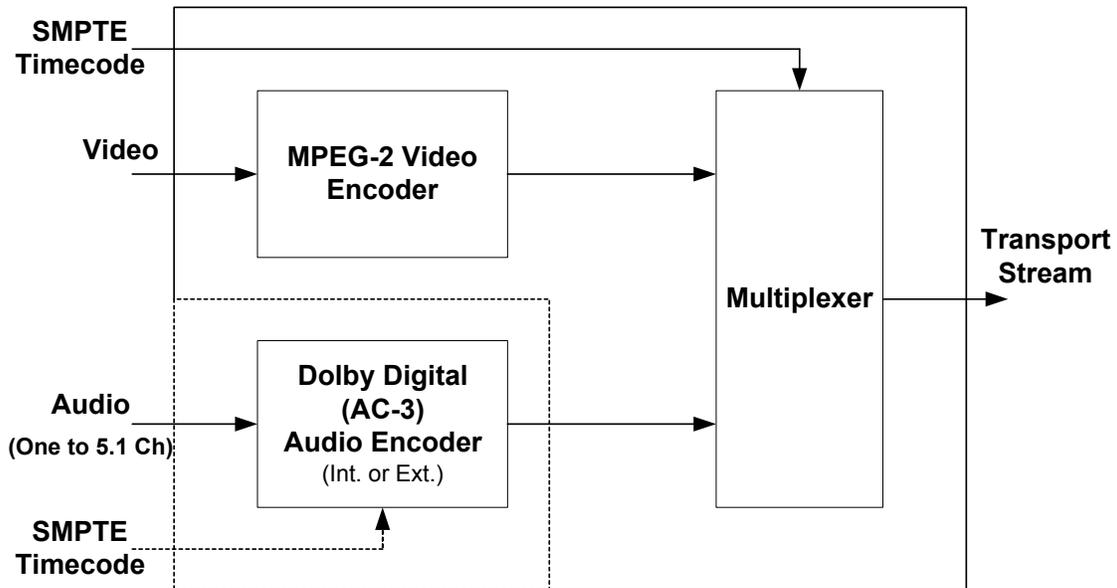


Fig. 1 – Simplified MPEG-2/Dolby Digital (AC-3) encoding system block diagram.

Aligning the MPEG-2 Encoding System

Video and audio encoding take some time to accomplish, and the multiplexer must know exactly how long. This delay depends on the manufacturer of the equipment, but the value is crucial for getting the PTS values correctly assigned. Many of the A/V Sync problems encountered thus far in the field can be attributed to these delays not being properly accounted for or just not set at all.

In practical terms, this simply means that if your transmission system uses an external Dolby Digital (AC-3) encoder, there is a known, fixed audio encoding latency that must be entered into the MPEG-2 encoding system. There is usually a setting called MPEG-2 Encoder Audio Delay, or possibly AC-3 Delay. Once set, it need not be changed unless either the audio or video encoder latency is reset.

In many cases, SMPTE timecode can be applied to the audio and video encoders and can be used by the multiplexer to calculate exact PTS values, thereby removing encoder delay as a source of error. All Dolby Digital encoders have the ability to accept external SMPTE timecode either as LTC or VITC signals.

Testing the MPEG-2 Encoding System

In its simplest terms, testing entails feeding typical A/V Sync test material such as beep/flash (audio pip with simultaneous video flash) to the encoder, capturing the resulting transport stream from the multiplexer, and using analysis software to determine compliance.

Although it is tempting, it is best to avoid using a consumer set top box to verify the performance of the MPEG-2 encoding system. There are commercially available tools that make the measurement easier and far more accurate by directly analyzing the MPEG-2 transport stream.

One such product, called “SyncCheck,” was manufactured by Interra Digital Video Technologies, Inc., but is unfortunately no longer available. This tool used software to demultiplex and decode the MPEG-2 and Dolby Digital (AC-3) streams, and displayed the decoded audio and video with both PTS and timecode values to permit fast and accurate sync testing.

The picture below was captured from SyncCheck and clearly shows the timing relationship between audio and video. Displayed is a sync test tape available from Sarnoff that contains 5.1 channels of audio and a frame of video that shows the active channels of audio. We are hopeful that such a useful software package will be developed and made available by another company soon.

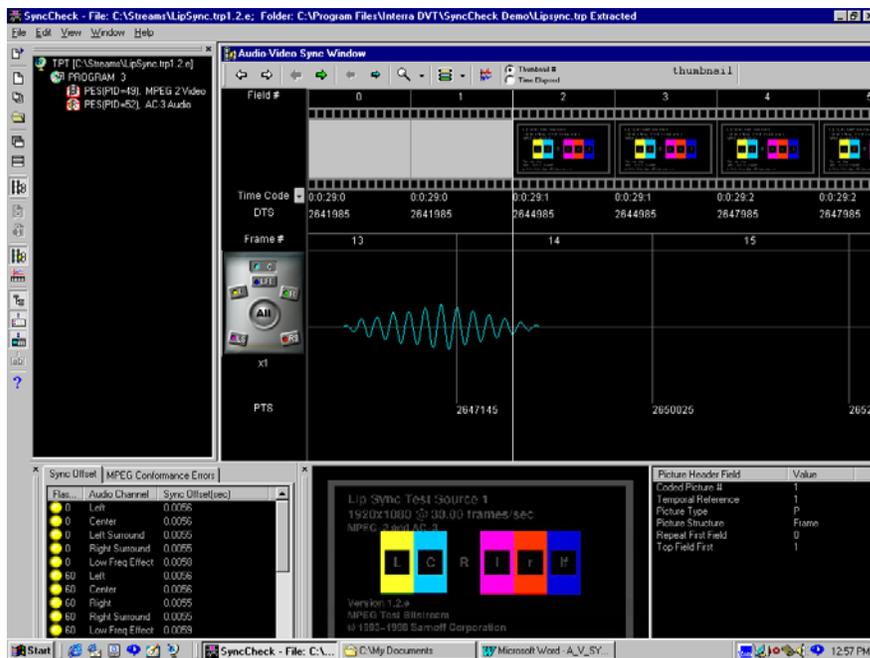


Fig. 2 – Interra SyncCheck main screen displaying results of encoded A/V sync test tape from Sarnoff. Note fractional sync offset between the Center channel and the start of the video frame. Individual channel values are shown in the lower left-hand box.

The latency of the Dolby Digital (AC-3) encoding algorithm is fixed regardless of the number of encoded audio channels. Therefore, a two-channel signal adequate to test the A/V sync relationship of an MPEG-2 encoded video signal and a Dolby Digital (AC-3) encoded audio signal. This means that a test tape can be easily created with a video flash and an audio beep, verified with an oscilloscope, and used as the source for testing with Sync Check.

Testing the MPEG-2 Decoder

Testing the MPEG-2 decoder is a very straightforward process. It requires that a reference transport stream be applied to the decoder under test. This transport stream is again a beep/flash type signal that has been encoded and verified for proper synchronization as described above. The audio and video outputs are then displayed on a dual trace oscilloscope and compared.

It is necessary to perform this testing with different video scanning formats and at different frame rates. This is due to the inclusion of video format converters after the MPEG-2 decoder that may respond differently to native rates than to rates that must be converted. Decoders with NTSC outputs should be carefully checked for A/V Sync “wander” at 24.0, 30.0, and 60.0 frame rates as 1/1000 video frames will be dropped to provide the 29.97fps output.

It is also important to test the decoder response to bitstream discontinuities as errors and splices are handled differently from one decoder to the next. **Figure 3** shows what a typical measurement might look like on an oscilloscope screen.

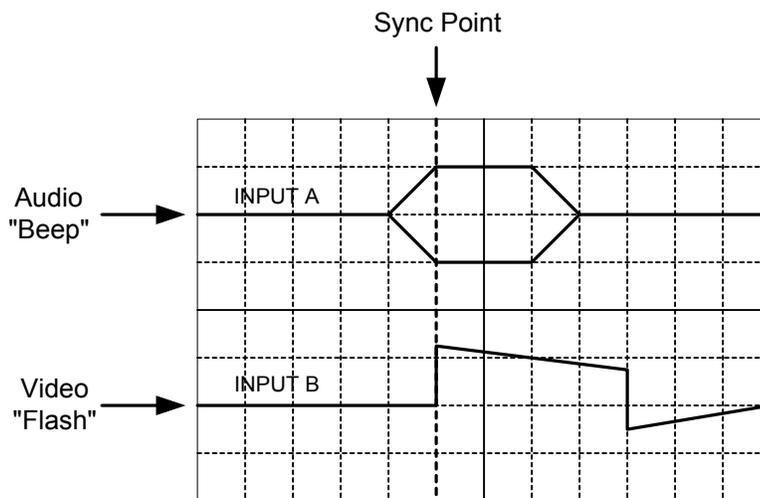


Fig. 3 – Audio and video outputs of decoder-under-test as would be displayed on a dual-trace oscilloscope.

In the display above, audio and video are exactly in sync. Note the ramp-up and ramp-down of the audio beep. This is related to the windowing function present in the Dolby Digital (AC-3) process. It can be seen that a proper measurement is made after the windowing, or when audio reaches maximum.

Conclusions and Recommendations

Hopefully by now it should be apparent that there are methods for creating programs that contain audio and video that are properly synchronized. It begins in the television plant with proper design of the facility. The simple rule of thumb is that audio and video should be exactly in sync before passing to the next process, especially before the MPEG encoder. There is equipment available for verification of sync, and equipment for correcting any errors. It is unwise to rely on the next destination of a program to correct for incorrect audio and video synchronization.

The MPEG system itself is inherently capable of guaranteeing synchronization between MPEG-2 video and Dolby Digital (AC-3) audio. There are commercially available tools to measure and correct this synchronization if necessary. Assuming correct sync at the input terminals to these encoders, timing calibration should be a one-time event.

In summation, careful planning and calibration are the keys to accurate and stable audio and video synchronization. Errors can creep in, but they are both measurable and correctable.

If you are having difficulty and need further assistance or guidance, please feel free to contact us. We would be happy to discuss your situation in detail.

Linear Acoustic Inc.
www.LinearAcoustic.com