

TAKING A BITE OUT OF LIP SYNC ERRORS Eliminating The Error Contribution From Production Switchers With Internal DVEs

Prepared By:

Chris Smith, Marketing Director

March 15, 2004

Introduction

The video and audio signals in our television system are being subjected to more and more steps of digital processing. Each step has the potential to add a different amount of delay to the video and audio, thereby introducing a lip sync error. Incorrect lip sync is a major concern to newscasters, advertisers, politicians and others who are trying to convey of trust, accuracy and sincerity to their audience. Studies have demonstrated that when lip sync errors are present, viewers perceive a message as less interesting, more unpleasant, less influential and less successful than the same message with proper lip sync¹.

Because light travels faster than sound, we are used to seeing events before we hear them – lightning before thunder, a puff of smoke before a cannon shot and so on. Therefore, to some extent, we can tolerate “late” audio. Unfortunately, as shown in Figure 1, even in a simple television system, the video is almost always delayed more than the audio, creating the unnatural situation of “early” audio. Any one contributor to the lip sync error may or may not be noticeable. But the cumulative error from the original acquisition point to the viewer can easily become both noticeable and objectionable.

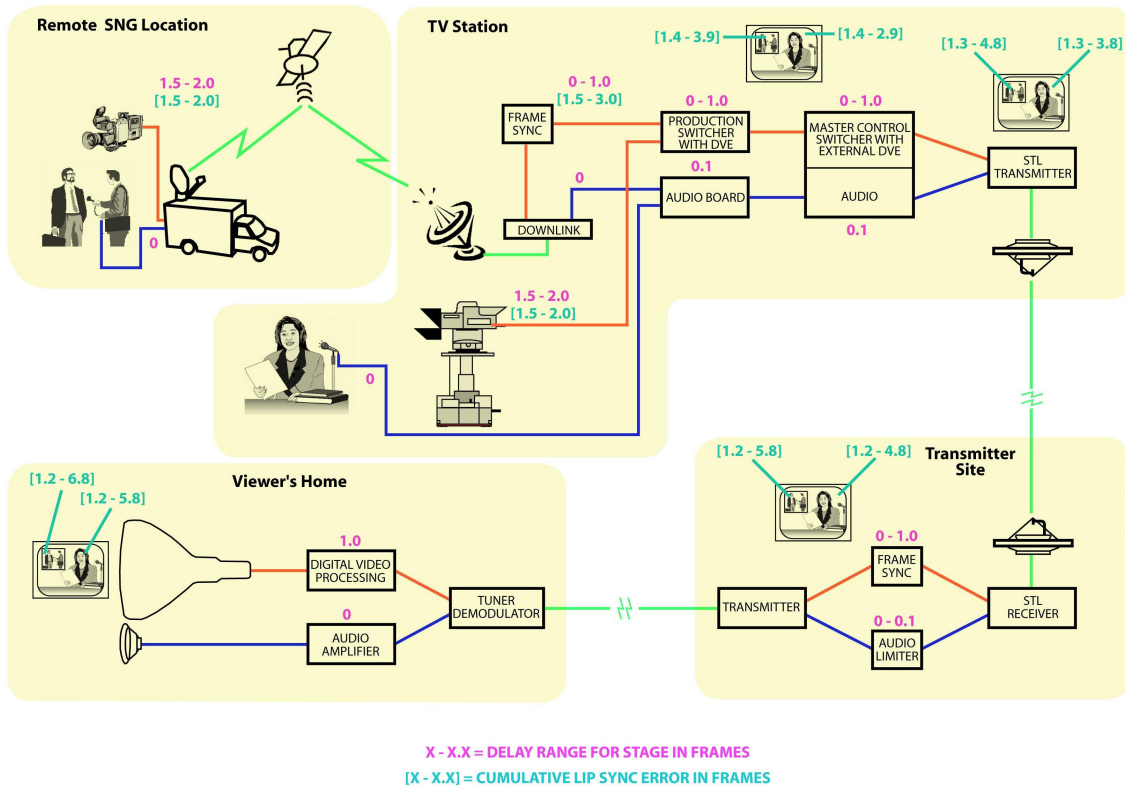
From CCD cameras, to frame synchronizers, production switchers, digital video effects, noise reducers, MPEG encoders and decoders, TVs with digital processing and so on, the video is delayed more than the audio. Worse yet, the amount of video delay frequently jumps by a frame or more as the operating mode changes, or as frames of video are dropped or repeated. So, using a fixed audio delay to “mop up” the errors is rarely a satisfactory solution.

Standards committees in various countries have studied the lip sync problem and have set guidelines for the maximum allowable errors. For the most part, these studies have determined that lip sync errors become noticeable if the audio is early by more than 25-35 milliseconds or late by more than 80-90 milliseconds. In June of 2003, the Advanced Television Systems Committee (ATSC) issued a finding² that stated “...at the inputs to the DTV encoding device...the sound program should never lead the video program by more than 15 milliseconds, and should never lag the video program by more than 45 milliseconds.” The finding continued “Pending [a finding on tolerances for system design], designers should strive for zero differential offset throughout the system.” In other words, it is important to eliminate or minimize the errors at each stage where they occur, instead of allowing them to accumulate.

Some Good News

Fortunately, the “worst case” condition in Figure 1 is less likely to present itself than a few years ago. Firstly, it is now quite common to install audio tracking delays (such as the Pixel Instruments AD-3000) alongside each video frame synchronizer, thereby eliminating at least one common source of variable lip sync errors. Secondly, newer master control switchers have an internal DVE for squeezeback operation rather than an external DVE. This allows the use of a constant insertion delay of 1 frame for both the video and the audio paths in all modes of operation.

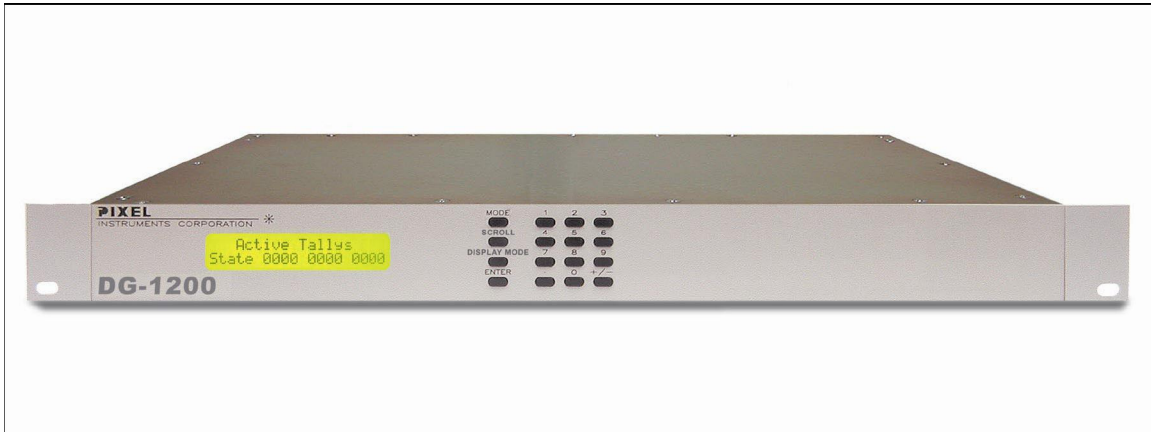
Figure 1: Cumulative Lip Sync Errors



The Production Switcher Lip Sync Problem

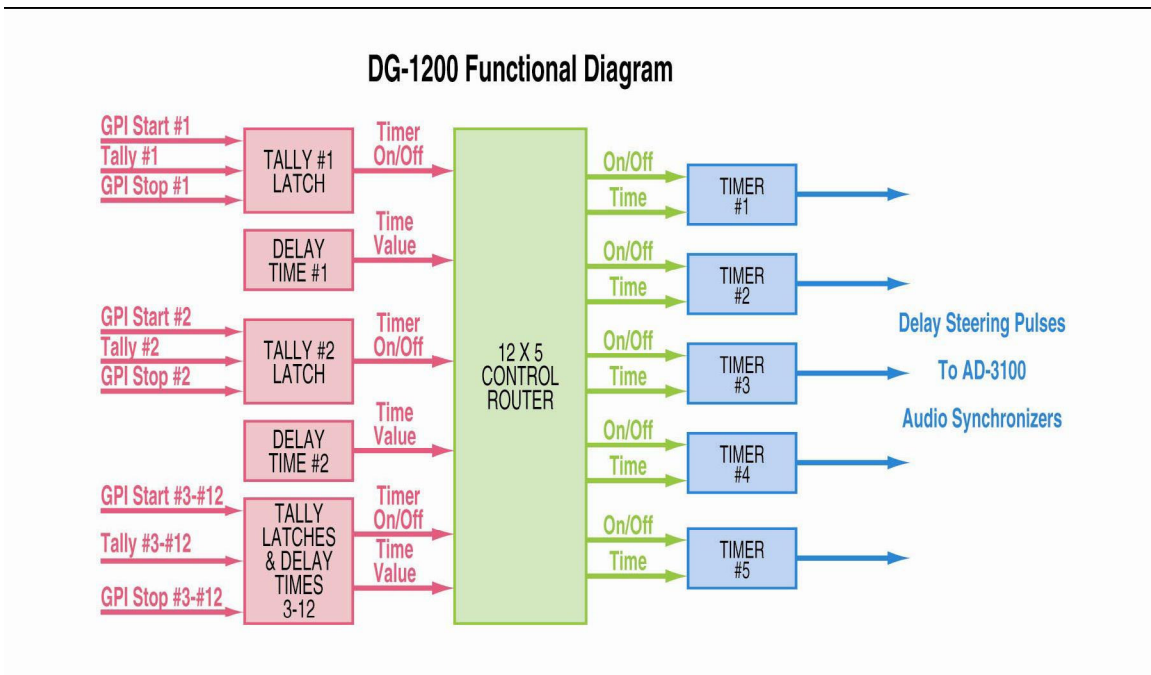
Since the 1970s, digital video effects processors (DVEs or transform engines) have been used to produce “over the shoulder”, “double box” and other multiple source composited effects. The video being transformed is delayed (usually by one or more frames) relative to the background video in the switcher. So, any time one or more DVE processors are on-air, the associated video sources will be delayed, resulting in a lip sync error. In the past, when the DVE processor was external to the switcher, a tally signal from the switcher could be used trigger the insertion of a compensating audio delay when the DVE is on-air. However, today’s production switchers are usually equipped with internal DVEs and a tally output is no longer available.

The Solution



Many of today's production switchers incorporate programmable timelines for the storage and recall of switcher configuration and effects. Typically a number of GPI and Tally contact closures can be stored in these timelines. The DG-1200 has been developed to interpret these GPI and tally outputs, generate the steering commands to control up to five audio synchronizers and automatically eliminate the lip sync errors. Based on the combination of effects being used in the switcher, the video delay is usually predictable. Therefore, the DG-1200 can be preset to provide the appropriate delay for each set of effects.

Figure 2: DG-1200 Functional Diagram



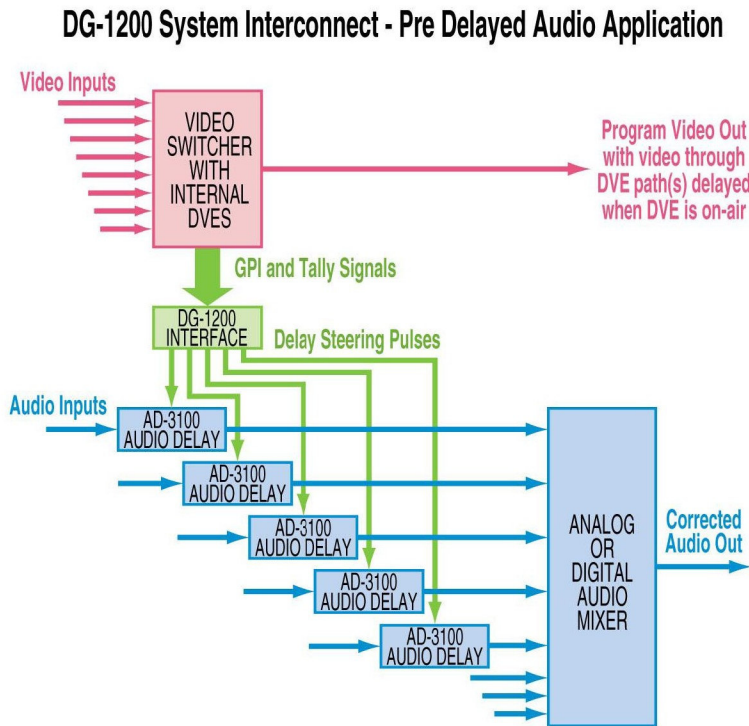
As shown in Figure 2, the DG-1200 has twelve input channels, each consisting of a GPI Start pulse, a GPI Stop pulse and a Tally line. Each input channel also has a linked delay time register with a user selectable value from 20 μ sec (nominally zero delay) up to 6.5 seconds, in increments of 100 μ sec. Delay times can be entered and displayed in milliseconds or in TV fields (NTSC or PAL). Input channels can be configured to respond to Tally only, GPIs only, or Tally gated by GPIs for maximum immunity to false delay insertion.

Any input channel and its time value can be routed to any of the five output timers and each timer can steer a separate AD-3100 Audio Synchronizer. The output timers can have different time values and can be turned on and off independently. Any timer can be controlled by more than one input. Let's say that one switcher effect needs a 1 frame audio delay and another effect needs a 2 frame audio delay. Input #1 (or any other input) can enable a 1 frame delay in Timer #3 (or any other timer) and the associated AD-3100. Any other input can be used to enable a 2 frame delay in the same timer.

Pre-Delayed Audio Application

The most comprehensive solution is to add AD-3100 Audio Synchronizers ahead of the audio mixer as shown in Figure 3. This configuration ensures that all sources contributing to the program output have the correct lip sync.

Figure 3

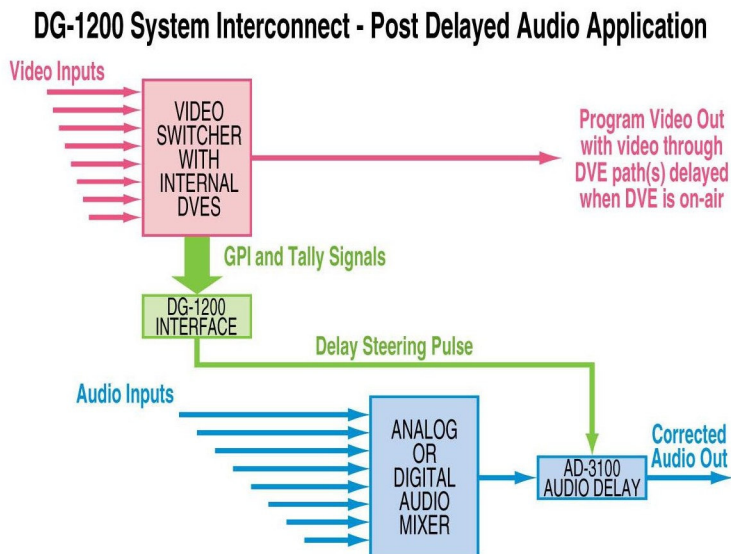


For applications that require more than 5 audio inputs to be delayed, this solution is scalable with additional DG-1200s and AD-3100s.

Post-Delayed Audio Application

In this simpler configuration shown in Figure 4, a single AD-3100 Audio Synchronizer is added at the output of the Audio Mixer. The amount of delay added to the audio path is chosen as a compromise for the sources contributing to the program output in any given effect.

Figure 4



For example, in a typical newscast over the shoulder shot, the studio anchor has zero video delay and the remote reporter (in the box) has 1 frame of video delay. Setting the AD-3100 delay to between 0 and 0.5 frame is the best compromise for both sources. The studio anchor's audio will be slightly late and the remote reporter's audio slightly early. Splitting the difference and choosing 0.5 frame delay is generally not the best choice since the early audio of the remote reporter is more noticeable than the delayed audio of the studio anchor. Adding the DG-1200 will reduce the residual lip sync errors compared to doing nothing at all.

Rapid Delay Change With Pitch Correction

The video delay of the DVE may be switched in and out of the program path several times in a relatively short time. Therefore, it is essential that the audio delay “catch up” quickly. The AD-3100 incorporates automatic pitch correction to allow rapid delay change without introducing undesirable artifacts such as pitch shifts, clicks and pops in the output.

Conventional audio synchronizers typically limit the rate of change of delay to around 0.5%. This means that for a 1 frame video delay change at the beginning of a program segment, the audio does not “catch up” until almost 10 seconds later. And another 10 second “catch up” period occurs at the end of the segment when the video delay reverts to normal. The AD-3100 has an adjustable rate of delay change of up to 25%. So, in our example of a one frame change in the video delay, the AD-3100 will “catch up” in just a few frames – well before the viewer will notice.

Conclusion

The combination of a tally/GPI interface (DG-1200) and a fast tracking audio synchronizer (AD-3100) provides a flexible cost effective solution to the lip sync errors introduced by production switchers and digital effects processors. It is also applicable to systems that use a master control switcher with external effects for squeezeback operation.

¹ Dr. Byron Reeves & Dave Voelker, research report Effects of Audio-Video Asynchrony on Viewer's Memory, Evaluation of Content and Detection Ability (1993)

² ATSC Implementation Subcommittee Finding, DOC.IS-191, 26 June 2003.