

# **A SHORT TUTORIAL ON LIP SYNC ERRORS,**

## **ABSTRACT**

This tutorial provides an overview of the causes of lip sync (audio to video synchronization) errors and their correction and cure. Examples of some of the more common sources of errors are given and correction methods using the Pixel AD 3000 and AD 3100 series audio synchronizers (tracking audio delays) are outlined.

Prepared by:

J. Carl Cooper  
Mirko Vojnovic  
Chris Smith

March 27, 2002 (updated August 12, 2004)

## THE SOURCES AND SOLUTIONS

One industry magazine reporter commented: "... how come some stations don't bother to correct the audio/video delay inequality? Sometimes when I watch Headline News I think I've somehow had my sound discriminator captured by a different channel's audio carrier. Hello? Anybody home? Compensating audio delay devices do exist. Honest!"<sup>1</sup>.

While compensating delays exist, the cure is not as easy as simply putting an audio delay in the audio feed to the transmitter.

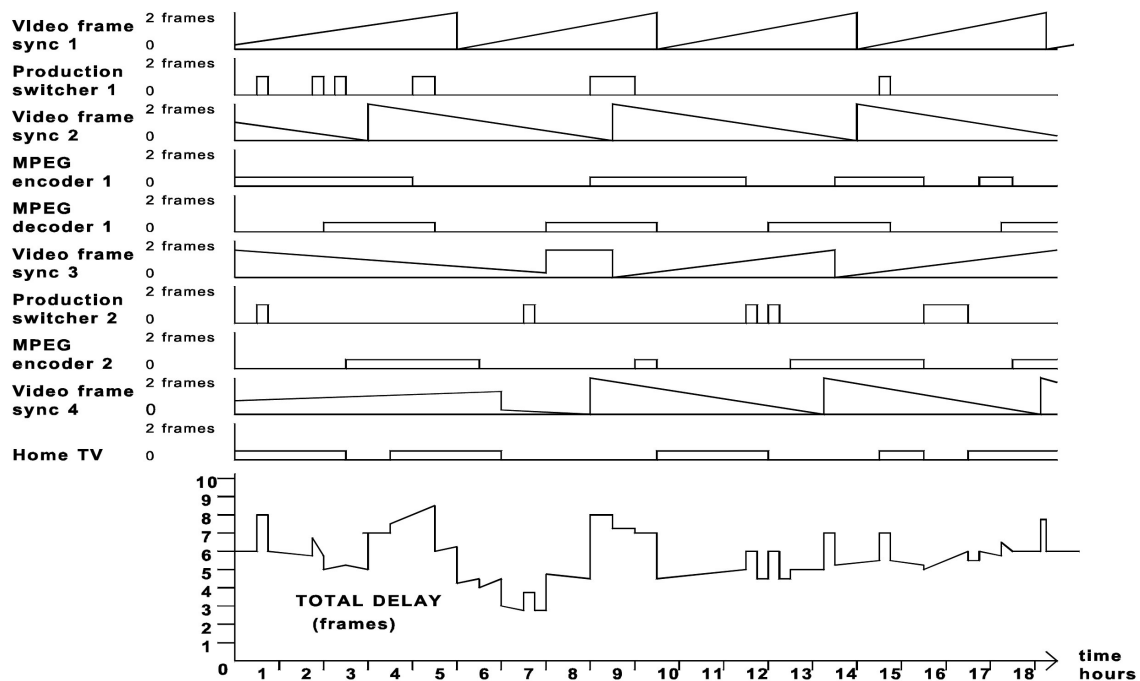
The causes of audio to video synchronization errors in video systems, commonly known as lip sync errors, are usually quite subtle. These errors are often the result of buildup of video delays at several locations, without provision for compensating audio delay. The problem is now so pervasive that it receives attention from virtually every segment of the TV industry.

Obviously the advertisers do not want to have their commercials aired with bad lip sync. Many are now monitoring for lip sync problems and asking for "free make-ups", when they find them. Bad lip sync is also a big concern to newscasters, reporters, politicians and others who are trying to convey a message of trust and sincerity to their audience. Without proper lip sync these people can be perceived by the viewers as less interesting, more unpleasant, less influential, and less successful than if the same person were viewed with proper lip sync<sup>2</sup>.

It is important to recognize that for even a simple system, video delays of several frames can occur. These video delays result in audio usually being advanced relative to video. Lately, as in case of some serial audio (eg. Dolby Digital) decoders, audio can actually be delayed in respect to video, which further complicates the problem.

Part of the mismatch actually occurs in the television camera due to delays in the CCD sensors, and part in the television signal production environment. Various video signal processing devices such as frame synchronizers, DVEs, A-D and D-A converters are typical video delay sources. It is important to note that many of these delays can instantly change video delay by throwing away or repeating frames of video to avoid video memory overflow. Jumps can also occur when the operating mode of these devices change.

A frequent problem with video delays takes place in MPEG encoding and decoding (as well as for other video delays) is the introduction of large quick video delay changes in the video. These delay changes are caused by instant changes in complexity and motion in the video image, and are often seen on network feeds. A typical system video delay chart for individual video delay components and the total delay is shown below.



Note that the error ranges from approximately 2.5 to 8.5 frames, with instant jumps of more than 3 frames at some instances.

From time to time, standards committees in various countries have addressed the problem of audio to video timing errors, and have set standards or guidelines for the maximum allowable amounts of these errors. For the most part, these standards or guidelines are in agreement that lip sync errors start to become noticeable if the audio leads the video by 25 ms to 35 ms, or if the audio lags the video by 80 ms to 90 ms. The smaller limit on advanced audio tends to support the theory that advanced audio is more annoying because it is an unnatural condition.

In June 2003, the Advanced Television Systems Committee (ATSC) issued a finding on this topic<sup>3</sup>. The ATSC finding states:

“One of the overarching goals of the DTV broadcasting system is to deliver audio and video in proper synchronization to the viewer.”

The finding continues:

“... at the inputs to the DTV encoding devices...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.”

and:

“Pending [a finding on tolerances for system design], designers should strive for zero differential offset throughout the system.”

In 1993, The International Telecommunication Union (ITU) was apparently thinking along the same lines. In its Draft New Recommendation [DOC. 11/59]<sup>4</sup> the ITU reported that errors of +20 ms (audio advanced) and -40 ms (audio delayed) were "detectable" and errors of +40 and -160 ms were "subjectively annoying". The draft recommendation suggested that an overall tolerance (for production, presentation, distribution and transmission) of +20 ms to -40 ms was appropriate.

Clearly, television facilities need to keep audio to video synchronization problems in mind. It is impractical to remove all of the video delays so the solution is to add compensating delays to the audio. Adding tracking audio delays, known as audio synchronizers, at every place where there is a substantial video delay is a good system solution to lip sync problems. In situations where an operator is already present (such as the master control room), providing an easy to adjust, fast tracking audio synchronizer allows the operator to correct upstream lip sync errors. Rapid corrections can be made on the air without introducing unwanted audio artifacts.

Typical applications are remote news feeds or network feeds that arrive at the station with a lip sync problem. This is usually due to the separate transmission paths that video and audio signals traverse (eg. video going through satellite uplink, and audio via land line).

It should be remembered that most video delays often change several frames in only a few seconds. The quick change in video delay requires the audio delay to quickly track greatly different values. Unfortunately the audio is continuous and there are no frames which can be repeated or deleted. Old style variable audio delay designs used memory address jumping and cause discontinuities in the audio data. Many products attempted to cover these pops, clicks and gaps, with only limited success. Even with the masking attempts, these variable delays still generate many undesirable artifacts in normal program audio, and are generally considered totally unsuitable for any broadcast quality audio.

The audio delay technology which has found the most successful solution uses first-in-first-out memories, or FIFOs. The FIFO has independent write and read to allow the reading to be faster or slower than the writing. The audio delay is controlled by causing the reading to catch up with the writing (to decrease delay) or to lag behind the writing (to increase the delay). This method works very well for making slowly changing delay adjustments in the order of 0.5% rate of change, and does not cause gaps, pops or clicks. Unfortunately the technique does create an audio pitch shift. This pitch error occurs because the ‘playback’ of the audio from the FIFO is not at the same speed as the ‘recording’ of the audio. There is a direct correspondence between the read/write rate difference and the pitch error. Normal people can hear pitch shifts of around 1% and people who have musical talent can hear on the order of 0.1%. If the rate of delay change is fast, then a noticeable pitch change is created.

With a small rate of delay change, the amount of time required to make large delay changes is correspondingly large, and the audio will be out of sync for a long time while the delay is catching up. For example, assuming a 0.5% rate of delay change is used, and an MPEG encoder generates a quick 6 frame change of video delay when a commercial is switched in. The audio delay will not be able to catch up to the new video delay for  $6 / 0.005$  or 1200 frames. That is 40 seconds and the lip sync will be way off for an entire 30 second commercial! Audio delays used for tracking video delays must have fast rates of change to keep up with the video, which leads to a noticeable pitch change in the audio.

In order to both minimize perceptible pitch shifts during normal small delay changes, and to allow rapid audio delay change after large video delay changes, it is necessary that the audio synchronizer incorporate a pitch correction circuit. With pitch correction capability, it is possible to make rapid delay changes with the pitch correction circuit reducing corresponding audio pitch artifacts to a level where the viewer will not notice them. The Pixel Instruments AD 3000 and AD 3100 audio synchronizers provide a high performance pitch shifting capability as part of the audio synchronizing function. As of the writing of this tutorial, they are the only audio synchronizers on the market with this capability.

<sup>1</sup> Mario Orazio, "The Masked Engineer", *TV Technology* 13.12 (December 1995): 5-10.

<sup>2</sup> Dr. Byron Reeves & Dave Voelker, research report Effects of Audio-Video Asynchrony on Viewer's Memory, Evaluation of Content and Detection Ability (1993).

<sup>3</sup> ATSC Implementation Subcommittee Finding, Doc.IS-191, 26 June 2003.

<sup>4</sup> International Telecommunication Union Document 11A/47-E, 13 October 1993.

<sup>5</sup> Randy Hoffner, "A/V Synchronization: How Bad is Bad?", *TV Technology*, 14 May, 2003.