

HIGH QUALITY VIDEO LEADS TO LOST REVENUES

ABSTRACT

This paper presents a tutorial on the problems of lower ratings caused by invisible audio to video synchronization errors and viewers subconscious perception of those errors. These errors often result from the use of high quality video instruments such as CCD cameras and DVEs. In this paper we will outline the problems caused by the errors, some of the more common sources of errors, and solutions to the error causes.

VIEWER PERCEPTION PROBLEMS

Ask any television engineer about audio to video synchronization errors and he will tell you that the result is visible "lip sync" errors. Unfortunately, what he won't tell you, and probably does not even know, is that small errors such as those which are not usually visible to even the master control director, can lead to lower ratings and consequently lost revenues.

Large, visible lip sync errors certainly can and do happen in today's systems, with the frequency of occurrence becoming a significant concern to advertisers and station management. However, small amounts of mistiming of audio and video which are often overlooked will cause a subconscious degradation of the program's entertainment quality as perceived by the home viewer.

The cause of this effect is believed to be the unnatural sound relationship which the television program presents. In our natural environment we are used to hearing audio slightly delayed with respect to video due to the slower speed of propagation of sound waves as compared to light. For example, we are used to hearing a racquet striking after we see the ball hit and hearing a commercial actor after we see them talking.

In today's television systems however, it is the video which is delayed thus causing the sound to arrive at the viewer's ears before the corresponding visual sensation. Viewing a television program with advanced audio is unnatural for the viewer, and is believed to cause subconscious stress. Psychological tests at Stanford University' demonstrate that viewers who watch television programs having advanced audio "evaluate people on television more negatively (e.g. less interesting, more unpleasant, less influential, more agitated, less successful)" than the same programs which were played with the audio in sync with the video. It was also discovered that this effect takes place with relatively small audio advances where the mere existence of an audio problem could not be detected by the viewers. The problem was also found to exist when the audio was delayed by more than the normally expected amount.

In addition to the negative perception of the program in the presence of advanced audio, there was also evidence the timing problem caused the test subject's memory of the negative aspects of the program to be remembered longer than normal. The worst

possible scenario takes place; the viewer gets a negative impression about the program, and also remembers it longer than a program which is properly presented. Obviously, such problems should cause a great deal of concern for station management.

Imagine if you will the effect this problem can have on the nightly newscast. Because of the audio sync problem, the audience perceives the newscaster, sportscaster, weathercaster, reporter, etc. as being less interesting or more unpleasant than they really are. The viewer remembers the negative feelings for a longer time than he would remember favorable impressions (1). The viewer responds to this negative feeling by turning to another station. Given the thousands of viewers involved, even if only a few percent are affected by the audio to video timing error, the results are lower ratings and consequently lower revenues and profits.

To recap, let us outline one very possible scenario:

- *Video is often delayed by processing instruments.*
- *Delayed video creates advanced audio Lip Sync error.*
- *Advanced audio causes subconscious viewer stress.*
- *Viewer stress causes the program to be unpleasant.*
- *Viewers turn to another channel/program.*
- *Viewers 'dislike' actors in commercials, do not buy product.*
- *Viewers 'doubt' commentators, don't accept message.*
- *Viewers 'wary' of politicians, don't vote for them.*
- *Advertisers aware of such problems are now watching for lip sync errors.*
- *Stations lose viewers/ratings due to viewer tune out.*
- *Reduced advertising revenue.*
- *Newscasters, actors, reporters lose viewer confidence.*
- *Reduced viewer confidence in station.*

Unfortunately, the cause of these timing errors is most often the use of high quality video equipment. As video quality demands increase, the amount of signal processing in the equipment increases, which in turn leads to video delays. When the video is delayed, it creates a corresponding relative audio advance, which if left uncorrected, can cause all of the evils pointed out above. A few of the typical video delays which are found today are explained below.

CCD CAMERA GENERATED VISION DELAYS

The wide use of cameras having CCD sensors aggravates the audio to video synchronization problem. All CCD sensors have an inherent visual delay mechanism. Depending on the sensor type and video processing used in the camera, the visual delay may be several fields (NTSC field = 16.7 ms, PAL field = 20 ms). In particular, the liberal use of digital frame store based image processing in high performance cameras is creating previously unknown vision delays of several fields, with a four field delay not being uncommon.

IMPROVED TEMPORAL RESOLUTION IN THE CCD

It would be worthwhile to mention the effect that variable shutter speeds, which are made possible by the use of CCDs, has on temporally sampling the image. At maximum exposure, corresponding to the slowest shutter speeds, the image is integrated over a long time, tending to blur any motion in the image. The blur makes it difficult for the viewer's brain to precisely distinguish such events as lip movement. This blurring (which was normal with tube based cameras) helps to mask the lip sync problem.

With the fast shutter speeds possible with CCDs, the sensor is in effect exposed for a relatively short time, which eliminates the motion blurring. In television systems, the ability to convey motion to the viewer increases dramatically with short exposure times. The shorter exposure time gives brighter and less blurred moving edges which result in the viewer's improved ability to perceive motion. As a consequence, the CCD camera's improved motion capability aggravates the corresponding increased video delay time and makes any audio to image timing mismatch easier for the viewer to detect, consciously or subconsciously.

CONSUMER SETS CONTRIBUTE TO THE PROBLEM

New large screen consumer TVs make a two pronged contribution to the synchronization error. First, it is believed that the larger screen size makes it easier for the viewer's brain to detect (and be disturbed by) an out of sync condition. Second, many large screen sets use their own digital processing in order to improve the visual performance of the set. Most common of these digital processing circuits are video noise reducers, progressive scan converters and Zoom processors, all of which add a frame or field of video delay. Unfortunately, very few TV sets compensate for the resulting video delay. Problem becomes even worse when audio is routed through separate multi-channel home theater audio amplifier.

VIDEO PROCESSING DELAYS ARE NOT CONSTANT

Video signals are often passed through digital video effects units, color correctors, noise reducers, frame synchronizers, compression equipment and a variety of other editing and image processing functions. As memory costs continue to decline, these devices increase in complexity, and many incorporate frame memory based processing functions which add delays which are switched in and out of the video path. Unlike the past where video

delays slowly drifted due to differing sync generator phases, the video delay in many of today's systems take instant jumps of one or more frames, as directors, editors and other operators select different processing modes. This situation is especially true of many current noise reduction and color correction products where extra frames of delay are added for each additional selected function. This instant change of delay length poses special challenges for the delay correction equipment and corresponding audio synchronizers which must keep up with these instant large changes in video delay.

SETTING PERFORMANCE STANDARDS

Recognizing the problems which can arise from small and undetectable timing errors, several committees have set standards or guidelines for audio to video synchronization errors. The Radio communication Study Groups of The International Telecommunication Union states (2) :

"Given the operating practices employed in the United States and the requirement that a single picture and sound service may reach the consumer in different forms and via different paths, the list of preferred points should be as noted above and the tolerances required at each of the points should be the same (+1field, -2 fields) with the understanding that these tolerances are absolute, are not accumulative, and apply to the overall system".

The International Telecommunication Union in the Draft New Recommendation [DOC. 11/59](3) reports that errors of and greater than +20 and -40 ms are "detectable" and errors of +40 and -160 ms are "subjectively annoying" (+ numbers indicate sound advanced with respect to video). The draft recommendation states:

A tighter tolerance on the range of values in the studio and production paths would be required to allow this [partitioning of tolerances]. The situation might look something like this:

- +20 ms -40 ms Overall tolerance*
- +10 ms -30 ms Production/presentation*
- +10 ms -10 ms Distribution/transmission*
- +2 ms -2 ms Per codec*

EIA/TIA-250-C standards(4) call for a +25 to -40 ms specification end to end for transmission facilities.

Very few station engineers are even aware of these standards, even fewer ever attempt to keep their stations operating within the standards. These standards are on the order of + 1 frame and most engineers would agree however these maximum permissible delays are well below the threshold of what would cause a noticeable error. Given the inherent video delays in today's consumer TVs and in high performance CCD cameras, very little additional delay can be tolerated in the rest of the system.

HALF HEARTED FIXES

Currently, some stations are attempting to fix their synchronization problems by inserting low cost fixed audio delays in their system. Unfortunately, this does not work since the video delays are constantly changing. The fixed audio delay merely serves to change the timing error from one where audio is always advanced with respect to video to one where audio may be advanced or delayed with respect to video. While this may reduce the easily noticed errors, for example by converting a 0 to -8 frame error to a +4 to -4 frame error, it does not cure the error. The only suitable cure is to delay the audio by the same amount as the video delay - not an easy problem when the video delay is constantly changing.

MEASURING THE VIDEO DELAY

Clearly, television facilities need to be designed with audio synchronization in mind. It is impractical to remove the offending video delays, so the only remaining solution is to ensure that the program audio receives the same delay as the associated video.

Part of the solution is to measure the video delay at each significant delaying device so that a corresponding audio delay can be inserted at that point. Several video synchronizer manufacturers have a digital delay output (DDO), which provide a current video delay value signal for use by a companion audio synchronizer. Additionally, video delay detectors are available for devices which do not provide DDO signals. The audio synchronizer receives the DDO signal and automatically delays the audio signal by a corresponding amount.

Delay detectors for measuring the video delay of devices without DDOs are also available.

Nevertheless, the problem with today's Delay detectors is that they insert certain reference signals into the programming material, thus affecting the content quality and most of them are to be used only in 'of the air' scenarios. Transparent and accurate measurement of varying 'in service' delays is the problem that has not yet been properly solved.

Pixel Instruments' LipTracker is the first product of its kind which measures A/V delays from the 'end users perspective', employing advanced machine vision and machine hearing heuristic algorithms, thus not requiring any alteration of the source material.

THE SECOND GENERATION AUDIO SYNCHRONIZERS

It was noted that the only currently viable solutions to the audio to video synchronization problem utilizes adjustable audio delays at some point in the system to delay the audio to match the delayed video. The adjustable audio delay remains a key element in television system designs, and second generation synchronizers are challenged with the problem of making adjustments to the delay which are imperceptible to the viewer.

As video delay values take jumps of one or more frames, the audio delay is required to take on the new, greatly different delay value without disrupting the audio. Old style

audio synchronizers often operated by dropping or repeating audio samples, and relied on slowly changing video delays to operate properly. The occasional sample manipulation usually went unnoticed by the home viewer. When faced with instant delay jumps of a frame or more, these old devices required several seconds or even minutes to catch up to the new delay values, creating noticeable distortion the whole time. Consequently, the audio would be both out of sync and noticeably degraded for the duration of the catch up. In today's systems where large jumps in delay are frequently made, this is unacceptable performance.

In order to overcome the problems inherent with sample manipulation, and more importantly to preserve the integrity of AES/EBU digital audio, it is necessary to have 1:1 correspondence between input and output samples in the audio synchronizer.

This correspondence can be achieved by varying the memory reading rate with respect to the storing rate to control the delay time. Varying the reading rate creates an annoying pitch change artifact however. In order to make the pitch change indistinguishable to the viewer, other types of old style audio synchronizers limit the differential rate between memory storing and reading to keep the associated audio pitch change very small. Unfortunately the small rate of change causes the amount of time to change delay settings to be correspondingly large.

The new generation of audio synchronizers (eg. Pixel Instruments AD3100, AD3000) minimize perceptible pitch shifts during delay changes with pitch correction circuits. The use of pitch correction allows rapid large delay changes and can maintain proper clock frequencies for correction of AES/EBU digital audio. With pitch correction, it is possible to make rapid delay changes, maintain proper clock frequency and remove any corresponding audio pitch artifacts so the change goes unnoticed by the viewer. While the amount of audio processing circuitry necessary to perform these functions satisfactorily causes significant cost increases with respect to simple fixed delays, it is currently the only viable solution to the problem. Considering the potential of lost ratings and corresponding lost revenues, however, the cost of the equipment to correct the audio synchronization problem is quite reasonable.

(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) International Telecommunication Union Document 10C/32-E, 11A/43-E, 11C/40E, CMTT-C/18-E 5 October 1993

(3) International Telecommunication Union Document 11A/47-E, 13 October 1993

(4) NAB Engineering Handbook, Television signal Transmission Standards (Washington, D.C.: National Association of Broadcasters), 621,