

A new generation of correction systems for audio-to-video synchronization errors has been developed for the TV broadcast industry. These products employ new techniques to solve a problem that has plagued the industry since its inception. However, recent production technology has opened even more opportunities for A-V sync problems.

Before describing the workings of this new hardware, it will be helpful to review the basic causes and effects of A-V sync problems including the results of some recent perceptual research.

Viewer perception problems

The most obvious result of audio-to-video mismatch is visible *lip-sync* error. This problem can and does occur in today's systems through a variety of means. The mistiming of audio and video always causes a subconscious degradation of a program's entertainment quality for the audience. This is particularly true when the audio is advanced with respect to the video. In natural environments, people are used to hearing audio slightly delayed with respect to video due to the slower propagation speed of sound vs. light. In today's TV systems, however, it is the *video* that is often delayed causing the sound to arrive before the corresponding visual image.

Beyond seeming unnatural, a TV program with such advanced audio (or delayed video) is believed to place the audience under subconscious stress. Psychological tests conducted at Stanford University have demonstrated that viewers who watched TV commercials with video lagging behind the audio responded to the people in the commercials more negatively (e.g., "less interesting, more unpleasant, less influential, more agitated, less successful") than the same commercials played with the audio in sync with the video. It was also discovered that this effect takes place with relatively small audio advances in which the actual identification of a lip-sync problem was detected by few viewers.

In addition to the negative perception of

Second-generation audio-to-video synchronization

the commercials in the presence of advanced audio, there was also evidence that test subjects' memory of the negative aspects of the commercial were retained longer in these cases. This presents the worst possible scenario to the advertiser: The viewer perceives the out-of-sync commercial in a bad light and remembers this negative appraisal longer than a commercial that is properly presented. Such problems can cause significant concern for TV advertisers.

CCD camera-generated vision delays

A-V sync errors are becoming more troublesome as TV technology progresses. The growing use of cameras with CCD sensors is one source of the problem. All CCD sensors have an inherent visual delay mechanism. The exact visual delay varies with the sensor type.

The inclusion of digital framestore-based image processing in newer CCD cameras exacerbates the problem, commonly creating video delays of up to four fields.

Variable temporal resolution in the CCD

Variable shutter speeds on CCD cameras also play a role in their temporal sampling of the image. At maximum exposure — a 1-frame shutter speed — the image is integrated over the entire frame, tending to blur any motion in the image and making it difficult for the viewer to distinguish

precise events, such as lip movement. This blurring was normal with tube-based cameras, which are continuously exposed to light.

With the fast shutter speeds of CCDs, the image is integrated over a relatively short time — for example, 100ms for a $1/10,000$ of a second exposure. The shorter exposure time gives brighter and less-blurred moving edges. This can result in a dramatic improvement of the viewer's ability to perceive detailed motion. Unfortunately, this improved perception also applies to the undesirable temporal artifacts of the signal. The CCD camera's effect on improved motion-perception makes any audio-to-image mismatch easier for the viewer to (consciously or subconsciously) detect.

Video-processing delays

Video signals are often passed through special-effects generators, color correctors, noise reducers, frame synchronizers and a variety of other editing and image-processing functions. As RAM costs continue to decline, these devices have increased in complexity, and many now incorporate frame-based processing functions that add delays that are frequently switched in and out of the signal path.

This presents a different problem from earlier days when video delays drifted slowly due to differing sync-generator phases. The video delay in many of today's systems can take *instantaneous* jumps of one or

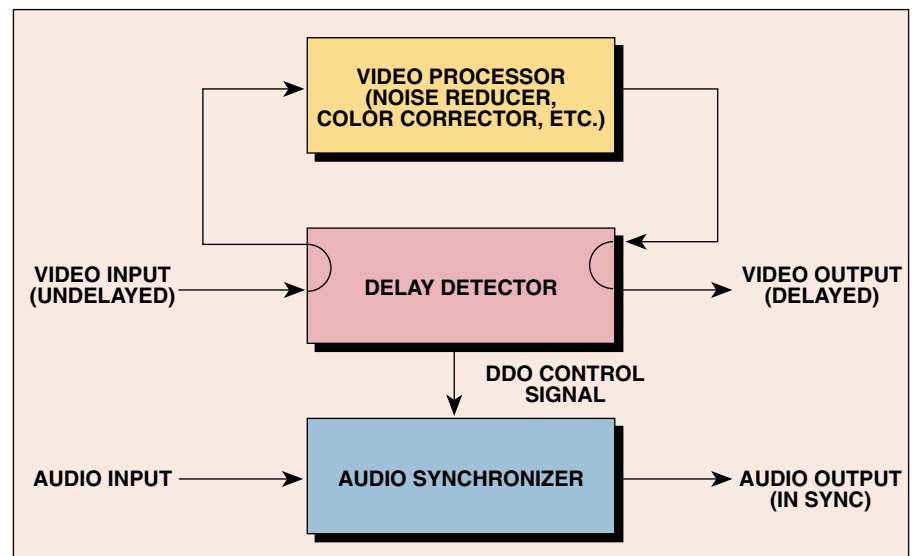


Figure 1. A delay detector (such as the Pixel Instruments DD2100) continuously measures the video delay caused by a video-processing device. The measured delay is sent as a digital delay output (DDO) control signal to a companion audio synchronizer (such as the Pixel Instruments AD2100 or AD3100), allowing it to continuously adjust its delay to keep audio in sync with video.

more frames as editors and other operators select different processing modes. This situation is especially true of many current noise-reduction and color-correction products in which extra frames of delay are added for each additional selected function. Such quick changes in delay time pose special challenges for audio synchronizers that must keep up with these relatively large and sudden variations in video delay.

Setting performance standards

Several committees have set standards or guidelines for A-V sync errors. For example, the Radiocommunications Sector of the International Telecommunications Union (ITU) specifies a tolerance of +1, -2 *fields* throughout a system. The “+” and “-” refer to the audio compared to the video (e.g., +1 field means the audio is one field ahead of the video). The ITU bases these recommendations on its subjective studies that indicate errors of greater than +20ms or -40ms are detectable and errors of +40ms and -160ms are subjectively annoying. (Again, “+” and “-” refer to sound’s arrival time with respect to video.)

The ITU’s recommendations further specify that these tolerances should be held absolute and not cumulative through a system, because signals can enter and exit a system at various places. In other words, these tolerances are not the net A-V sync effect on a signal passed through an entire facility, but instead are the tolerances observed at every point throughout the facility. An ITU draft recommendation calls for a *partitioning of tolerances* throughout a facility as shown in Table 1.

EIA/TIA-250-C standards call for a similar +25ms to -40ms specification for transmission facilities (end-to-end). Given the inherent video delays in CCD cameras, little additional delay can be tolerated in the rest of the system.

Measuring the video delay

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

Part of the solution required to satisfy the ITU recommendations involves measuring the video delay at each significant delaying device so that a corresponding audio delay can be inserted at that point. Several video synchronizer manufacturers provide a *digital delay output* (DDO) that indicates a current video delay-value for use by a companion audio synchronizer. Alternatively, video delay detectors are available for devices that do not provide DDO signals. The audio synchronizer receives the DDO sig-

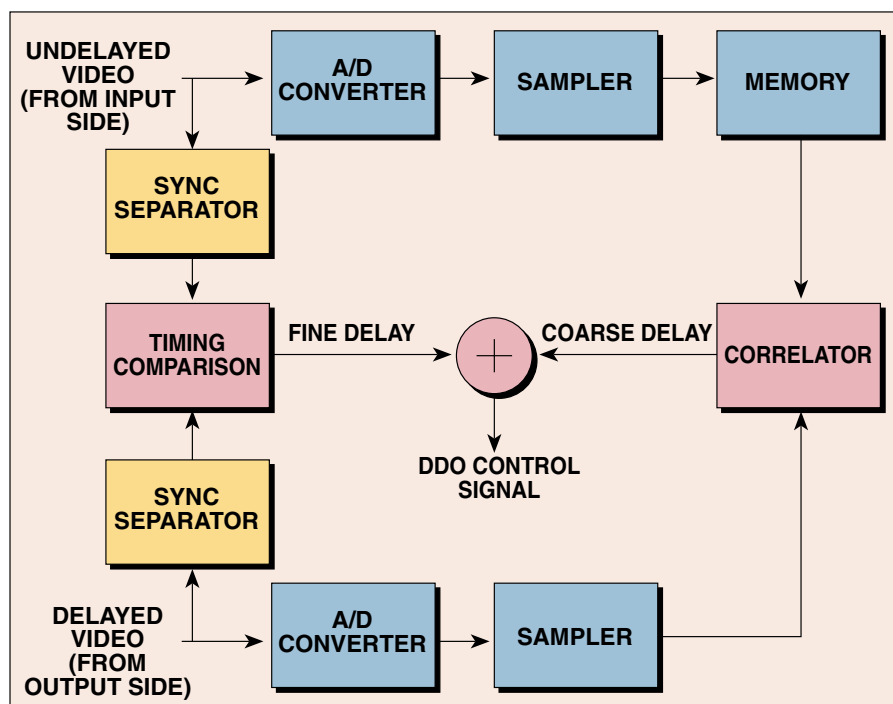


Figure 2. Basic block diagram of the Pixel Instruments DD2100 delay detector.

nal and automatically delays the audio signal by a corresponding amount.

Delay detectors for video devices without DDOs operate as follows: The detector has two video inputs. One is fed by tapping the video signal just as it enters the device in question, and the other is fed by the device’s output. (See Figure 1.) The delay detector stores a given video frame from the upstream input and then compares all the frames coming out of the device to the stored frame. By counting the number of frames that pass until the stored frame is output, the video delay through the device is determined. Comparing the phase of the device’s input and output vertical sync signals allows the delay to be measured with subframe precision. (See Figure 2.)

These detectors are easy to add to an existing system requiring only that input and output video be looped through their inputs. They provide a DDO signal that can be used by an audio synchronizer to make appropriate corrections.

The second-generation audio synchronizer

All currently viable solutions to the A-V sync problem use adjustable audio delays at some point in the system to delay the audio to match the delayed video. Because current video processing can create large and abrupt video delay changes, today’s audio synchronizers are challenged by having to make dynamic delay adjustments that are imperceptible to the audience.

Older audio delays often worked by drop-

ping or repeating audio samples and relied on slowly changing video delays to operate properly. The occasional sample manipulation that these devices performed usually went by unnoticed. However, when faced with instant delay jumps of a frame or more, these devices require several seconds or even minutes to attain the new delay values with the sample manipulation, often creating noticeable distortion during the entire adjustment period. This causes the audio to be out of sync and noticeably degraded for an extended time. In systems where large jumps in delay are frequently made, this produces unacceptable performance.

An additional problem that has become more important is the effect of such sample manipulation on audio outputs that remain in the digital domain (typically using the AES/EBU interconnection standard).

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

This implies that the audio delay memory must store every incoming audio sample and retrieve every stored audio sample only once. In order to accomplish this, the memory must have completely decoupled and asynchronous storage (*write*) and retrieval (*read*) functions. By varying the read rate with respect to the write rate, the delay time can be controlled either decreasing delay by causing the retrieval to catch

STAINLESS
2C
FILM ENCLOSED
PLEASE REMOVE HOLDING LINES

AUDIO LEADS VIDEO BY	AUDIO LAGS VIDEO BY	MAXIMUM TOLERANCE PER SECTOR
+20ms	-40ms	Overall tolerance
+10ms	-30ms	Production/presentation sector
+10ms	-10ms	Distribution/transmission sector
+2ms	-2ms	Per codec (within sectors)

Table 1. Partitioning of tolerances for audio-to-video synchronization errors as proposed by the ITU. (Source: ITU Doc. 11A/47-E, October 1993.)

up with the storing or increasing delay by having the retrieval lag behind the storing. In digital audio systems, this must be performed with the independent requirement of maintaining the word-clock rate at the correct frequency.

Varying the read rate with respect to the write rate can create an audible pitch-change artifact, and it also requires re-clocking to maintain the proper clock rate for AES/EBU output.

The only comprehensive solution is the inclusion of an integrated pitch-correction circuit in the synchronizer.

One way to make the pitch change indistinguishable to the viewer is to limit the differential rate between write and read rates, thus keeping the audio pitch change extremely small. Unfortunately, this results in making the amount of time required to change delay settings correspondingly large.


Another possible approach involves an adaptive process, whereby the frequency and level of the audio signal is taken into account. Larger read/write rate ratios can be tolerated if little or no high-frequency audio is present or if there are periods of silence. This does not provide a consistently significant improvement, however, and frequently is of no advantage for any program material having a musical background.

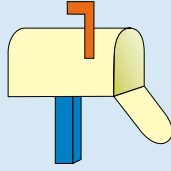
The only viable, comprehensive solution is the inclusion of an integrated *pitch-correction* circuit in the audio synchronizer. With such pitch correction, it is possible for the synchronizer to make

rapid audio delay changes and maintain proper digital output word-clock rate while the pitch-correction circuit adaptively reduces any corresponding audio pitch artifacts to a level that will be unnoticed by the audience. Given current digital signal processing (DSP) technology, this functionality is possible and affordable.

One commercial product that incorporates pitch correction is the AD-3100 manufactured by Pixel Instruments. This device has selectable analog and AES/EBU digital inputs along with simultaneous analog and digital outputs. It receives a DDO signal from a video device and adjusts the read rate of the internal memory to increase or decrease the delay while at the same time providing DSP pitch correction to maintain proper pitch and output sample rate. In this device, multiframe delay changes can be made in a matter of milliseconds without introducing noticeable audio artifacts or losing A-V or digital synchronization. ■

J. Carl Cooper is director of engineering at Pixel Instruments, Los Gatos, CA.

 For more information on Pixel Instruments audio-to-video synchronization equipment, circle (100) on Reply Card.

 Internet: be@intertec.com
CompuServe: 74672,3124
FAXback: 913-967-1905