

COMMERCIAL LIP SYNC PROBLEMS

In today's digital (and hybrid digital/analog) broadcasting facilities the opportunities for lip sync errors continue to multiply. From CCD cameras, to frame synchronizers, production switchers, digital video effects, noise reducers, MPEG encoders and decoders and so on, the video is typically delayed more than the audio. Worse yet, the amount of video delay frequently jumps by a frame or more as frames of video are dropped or repeated. Or, in other cases, digital audio surround sound processing can cause the audio to be delayed by several frames relative to the video.

If these lip sync problems occur during paid advertisement programming it may have adverse effect on TV station's business. When the advertisers see the problem, they may refuse to pay for the commercial, or demand a makegood.

Since the occurrence of lip sync errors is inevitable we are faced with two issues – how to measure the errors and how to correct them quickly and transparently.

Measuring Lip Sync Errors

In the last few years there has been little change in practical measurement methods for lip sync errors.

- The human eye and ear method works. However, it does not provide a precise measurement of the number of frames error and does not identify whether the problem is occurring at one stage or is the sum of several errors occurring at multiple stages. Also, it is labor intensive and does not trigger an automatic alarm every time an error occurs.
- Another method is to insert “watermarks”. One example is Tektronix AVDC100 in which the audio envelope is included in the video as least significant bits (LSBs). This method at the first glance looks very elegant in its conception and execution, but compression systems, when heavily compressing, may lose the data of the watermark signal because the LSBs are discarded as “insignificant, thus unnecessary”. Watermarking may be suitable in controlled environments, but as soon as the signal is passed between two non-compatible processing plants the watermark may be lost or damaged, and accordingly, the synchronization information is lost. Similar problems are faced by all other “watermarking” schemes.
- A possible method is to have VITC (Vertical Interval Time Code) with the video and LTC (Linear Time Code) with the audio. This works, for example with Dolby E, which includes LTC. Also DATC (Digital Audio Time Code) can be used with AES audio to convey LTC, but it requires additional equipment in the transmission path at one location for DATC insertion at the source, with accompanying decoder on a DA output at the destination.

The difference in seconds and frames between the VITC and the LTC could be used to monitor lipsync. This is assuming that the video is being conveyed at the same rate as VITC, but that may not be the case unless the compression system is designed to halt or slow the VITC and/or DATC numbers whenever the video is halted (frozen) or slowed down. So this method is dependent on compression system implementation, and as these simultaneous

VITC-video and LTC-audio rates are not part of the compression system specifications, there is no compulsion to implement it. At the same time, insertion of DATC is just one more form of “watermarking” and therefore it is prone to become lost down the stream.

- The LipTracker lip sync analyzer is a non-invasive measurement tool for in-service lip sync analysis. After detecting a face in the video, LipTracker compares selected sounds in the audio with the mouth shapes that create them in the video. The relative timing of these sounds and corresponding mouth shapes (called Mutual Events or MuEvs) is analyzed to produce a measurement of the lip sync error. The sounds and mouth shapes that are used for MuEv analysis are commonly found in the natural speech patterns of many languages.

Numeric and graphic displays of the current audio offset are updated periodically until the current face is lost or a new face is detected. A history graph charts the most recent error profile and event logging saves the results for scene by scene analysis. An Audio Offset Status indicator provides a visual warning of the current offset.

This unique approach of analyzing real time video and audio content does not require the insertion of cues, codes or watermarks into the program material. Therefore, LipTracker can be used at any point in the transmission path.

Note: Detailed information regarding LipTracker is available at www.pixelinstruments.tv .

Correcting Lip Sync Errors

Even in facilities that use tracking audio delays, the results may be unsatisfactory. Some competitive audio synchronizers cannot track delay changes quickly, so if the lip sync is wrong at the start of a commercial it will probably be off for the entire commercial. Or, unwanted audio artifacts may be introduced. The delay change problem stems from the fact that audio is continuous; you cannot simply drop or repeat a frame as you do with video in order to make an instant delay change. In one of our competitor's products, audio samples are dropped or repeated to make delay changes. The manipulation of samples causes clicks and pops in the audio during the period of time the delay is changing. Advertisers are not going to be happy with clicks and pops in their commercials.

Another competitor simply limits the rate of change of their audio delay. This eliminates the clicks, pops and distortion; and keeps the pitch shift very low. Limiting the rate of change means it often takes two or three minutes for the audio delay to catch up with the video after a one or two frame delay change. By the time the audio catches up the commercial is over. This also makes the advertisers unhappy.

The Pixel AD3000 and AD3100 audio synchronizers incorporates pitch shifting and fast track technology to avoid these problems. Our synchronizers can make a ½ second delay change in less than 2 seconds, and do it invisibly. Most experienced television engineers cannot hear any delay change artifacts from these rapid corrections, even when told they are going to happen in the next minute.

The fast track technology of our AD3000 and AD3100 is easily demonstrated. Simply program the synchronizer to a fixed 0 second delay, run any program audio through the unit, and have the engineer listen to the audio but face away from the AD3000. Next, tell the engineer that you are going to change the delay in the next two minutes and quietly program the delay to .5 second. The most that anyone can possibly hear is a very slight tempo change in music, and that usually takes a musician. Most people, even experienced engineers, hear absolutely no artifacts. Repeat the demo and let them watch the front panel so they can see how fast the delay changes. For fun, set the delay to 2 seconds and let them listen to the audio during the delay change.

Try the same demo with any of our competitors' products and one or more of three audio artifacts will be immediately apparent. There will be clicks, pops or distortion, there may be a pitch change, and it may take a very long time for the audio to change. With all of our competition, there is a tradeoff between pitch change, distortion and time that it takes to change the delay. Try changing the delay from 0 to 2 seconds and then back to zero, noting how long it takes or how noticeable the artifacts sound. Our AD3000 family of products run circles around the competition in all respects. In any case, it only takes one bad commercial to make the advertisers unhappy. The loss of revenue from one commercial may very well be more than the cost of one of our audio synchronizers with pitch correction.

Another important point is that viewers find speakers who do not have proper lip sync as less interesting, less trustworthy, etc. as compared to those where the lip sync is correct⁽¹⁾. This of course is a big concern to newscasters, reporters, politicians and others who are trying to convey a message of trust and sincerity to their audience.

Of course stations are concerned with keeping their news ratings up and they don't want their viewers thinking that the newscasters are not interesting. This has also become an issue to a number of federal politicians, and Congress is looking to upgrade the entire C-SPAN system to guarantee that lip sync is always correct. After all, what politician wants to be perceived as a crook simply because his speech gets aired with a noticeable lip sync error.

1. Dr. Byron Reeves and Clifford Nass, *The Media Equation* (Stanford, California: Stanford University Center for the Study of Language and Information), 211-218