



Audio/Video Synchronization

Prepared by Simply Labs, LLC
www.simplylabs.com

I. Lip Sync Defined

When entertainment content is decoded and rendered on Consumer Electronic (CE) devices, the timing of rendering the video portion of the signal may deviate from the timing of rendering the audio signal. The resultant timing differential is often referred to as a "lip sync" error, since it is most obviously apparent to a viewer when the content contains a representation of a person speaking. In a digital television, the video processing usually takes more time than the audio processing. Because of this, synchronization of video and audio can become an issue, creating an effect similar to a badly dubbed movie, where the audio and video don't match up and the sound of the spoken words is no longer in "sync" with the speaker's lip movement.

With all the emerging technologies coming to market, it helps to understand exactly what a specific term means. HDMI version 1.3 includes a Lip Sync feature, which allows the audio processing times in devices to be adjusted automatically to compensate for errors in audio/video timing. The initial implementations of this functionality will be in A/V receivers, but it is likely to appear in DVD players and many other CE devices in the future. Reports from manufacturers indicate that this function is very popular and will be widely implemented.

The HDMI standard requires manufacturers to disclose specific HDMI features enabled in a product. The idea is to provide consumers with the necessary descriptive information they need to understand enabled features that exploit certain capabilities of HDMI, such as Lip Sync. For each feature, the guidelines specify a minimum level of functionality that must be met by the device in order to use the terminology.

While HDMI LLC Authorized Testing Centers (HDMI-ATCs) test for electrical parametric and protocol compliance against the HDMI specification, there is a need to build upon this basic interface testing with additional performance testing programs designed to simplify consumer purchase decisions and enhance the high definition entertainment experience. There are no HDMI-ATC system level Lip Sync performance compliance specifications, or test tools designed to ensure accurate Lip Sync feature delivery. There is no "timing conformance" specification that must be demonstrated to any authority in order to build a compliant product.

There is an increasing awareness in both broadcast engineering and the CE industry that audio-video synchronization errors, usually seen as problems with lip sync, are occurring more frequently and often with greater magnitude. With the advent of digital processing in CE devices, the issue has become critical. Some CE manufacturers deny there is a problem, believing the audio/video asynchronies in their units to be imperceptible. Knowing how to measure audio/video delays and compensate for them has become increasingly important, as described in this White Paper.

II. Is it Important?

Lip Sync is very important to consumers and the display industry since newer technologies have created a noticeable delay between the processing of video signals and the processing of audio signals. Lip Sync correction features take into account processing delays, so that both signals can be synchronized and presented to the viewer together. This greatly improves the entertainment experience for the viewer.

Generally, it is desirable to minimize or eliminate lip sync errors because they detract from the consumer entertainment experience. The lack of lip sync correction is of particular concern in certain types of content, such as product commercials and political candidates' statements. See the report "*Effects of Audio-Video Asynchrony on Viewer's Memory, Evaluation of Content and Detection Ability*" by Reeves and Voelker for more information (a non-copyrighted [PDF](http://www.pixelinstruments.tv/pdf/Articles/Effects%20of%20Audio-Video%20Asynchrony.PDF) is available at <http://www.pixelinstruments.tv/pdf/Articles/Effects%20of%20Audio-Video%20Asynchrony.PDF>).

Human studies conducted for sensitivity to audio/video asynchronies have shown that a drift where the audio arrives late is not as annoying as when the audio arrives early. In fact, even a few frames of early audio can quickly be detected by the viewer. The characterization of sensitivity to the alignment of sound and picture includes early work at Bell Laboratories.

The extent to which these asynchronies can be tolerated by a consumer is dependent upon human perceptual limits as well as personal taste. Steinmetz and Engler conducted user studies [R. Steinmetz and C. Engler, "Human Perception of Media Synchronization," *Technical Report 43.9310*, IBM European Networking Center, Heidelberg.], and they report several figures of merit for quantifying tolerable audio/video asynchrony limits.

In 1998, ITU-R published BT.1359, recommending the relative timing of sound and vision for broadcasting. Studies by the ITU and others have suggested that thresholds of timing for viewer detection are about +45ms to -125ms, and the thresholds of acceptability are about +90ms to -185ms. In addition, the ATSC Implementation Subcommittee IS-191 has found that under all operational situations, the sound program should never lead the video program by more than 15ms and should never lag the video program by more than 45ms ±15ms.

When viewers encounter difficulties such as lip sync errors, blocking or black screens, they turn to another channel. Therefore, it is imperative that television engineers find and fix network, encoding, and transmission problems before their viewers become aware of them.

III. Problem Origins

Lip sync errors are becoming a significant problem in the digital television industry because of the large amounts of video signal processing used in television production, television broadcasting, and pixilated television displays such as LCD, DLP and plasma panels. Audio and video synchronization problems occur because video processing is more intensive than audio processing. Because of this, the audio is ready for playback before the video, and if audio is not delayed, what viewers hear will not match what they see on the screen.

Anywhere video is processed, there will be a delay. Video processing filters, format conversion, compression — all of these will add delay, perhaps as little as a few pixels or one line of video, or perhaps as much as many frames of video. Although faster processors and clever algorithms can minimize these delays, they can never completely eliminate them. Ignore the delays, and you have audio and video out of sync. On the display side, video processing delays become significant for LCD and plasma display panels (PDPs), where memory-based video-processing algorithms, as well as panel response times, can cause a delay of more than 100ms.

Compressed and broadcast video brings yet another difficulty in the form of variable delays. Since the amount of compression varies with video material, the instantaneous compressed bit rate (bits per frame, for instance) will vary as well. In order to use bandwidth efficiently, the rate needs to be smoothed to an overall constant bit rate, and that means that the delay will vary.

Yet another origin of synchronization failure is when different audio/video system components (or even STB/TV tuner channels for that matter) are used in the chain. In other words, the audio/video delay can actually jump to a different value when a new device is inserted within the stream.

Finally, audio synchronization issues can also arise from wireless multi-channel speaker applications. Because of the inherent processing delays of wireless transmission, it takes more time for wireless transmitted channels to output audio than non-transmitted audio channels. The non-transmitted audio channels, therefore, need an additional delay to synchronize them with transmitted channels. As television applications become more and more sophisticated, and more wireless components are being adopted, a real need has developed for Lip Sync correction solutions.

IV. Industry Activities

It is becoming apparent that end-to-end solutions are needed, and several trade groups are actively studying the problem. SMPTE has created, within the S22 Committee on Television Systems Technology, an Ad Hoc Group on Lip Sync issues to address the problem and produce guidelines. Work on a coordinated studio-centric solution will probably include problem assessment, current practices, control signals, and potential solutions.

Television industry standards organizations have also become involved in setting standards for audio/video sync errors. See for example ATSC Document IS-191 (http://www.atsc.org/standards/is_191.pdf). Although the AC-3 digital audio standard is mature, implementations have varied, in particular with regard to lip sync. The ATSC Technology and Standards Group on Video and Audio Coding (TSG/S6) has been directed to look into these issues, and has established two working groups to gather implementation data and report back with recommendations.

In Canada, World Broadcasting Unions International Satellite Operations Group (WBU-ISOG) has conducted tests on satellite encoders and decoders. An EBU audio group has performed tests for SDTV receivers. In Japan, the Japan Electronics and Information Technology Industries Association (JEITA) IEC-TC100 has started investigations on TV receiving devices.

V. Lip-Synch Detection & Reduction Methods Survey

Aside from a handful of proprietary solutions, no standard solution has yet been proposed. The causes are many, but well defined lip sync detection solutions are hard to find. It is difficult for humans to easily determine how much video-to-audio delay or advance is present.

Existing CE methods for reducing the timing error between the audio and video portions of content have included manual adjustment based on a delay factor determined by observation by the consumer, or automatic adjustment as in HDMI 1.3 based on a previously estimated delay factor. The disadvantage of a manual measurement and adjustment is that it is based on a human-perceived delay.

The method of automatically delaying the audio by an estimated factor, based on the expected delay in the video signal during processing, is also an imperfect solution since the audio and video signals may be routed through a number of devices, and may undergo a number of processing steps. Each additional device or step can impact the ultimate lip sync error. Some HDMI 1.3 A/V Receivers are incorporating an audio synchronizer, which can be used to correct or maintain proper audio/video sync. In order to correct audio/video sync problems the HDTV outputs timing delay information, propagating the amount of delay the video signal experiences. The audio synchronizer receives the delay information and in response delays the audio by an equivalent amount, thereby theoretically maintaining proper synchronization.

The actual delay needed for synchronization will depend on the type of audio and video signals, and the current video mode. Video delays can frequently and rapidly change by large amounts, requiring the audio synchronizer to achieve corresponding changes in the audio delay without introducing audio artifacts such as pops, clicks, gaps, or pitch errors. Unfortunately, the video delays frequently make quick and large changes. In order to maintain proper audio/video sync, the audio delay needs to track these video delay

changes. This rapidly changing information is typically not conveyed in real-time with today's products.

For broadcast applications, broadcast stations are investigating the use of set-top boxes (STBs) and video monitors to confirm that they are correctly in sync. Digital broadcasting, however, introduces another element to the monitoring equation — software. Every digital STB has software running on it. Depending on the implementation of the software in a specific STB, the receiver may react differently to a specific channel within the signal stream. Therefore, problems that affect users of one type of STB may not be visible to users of another STB brand, or even a later model from the same manufacturer.

Various audio/video synchronization technologies for broadcast currently exist that can analyze, measure and correct lip sync error. One measurement system uses a special test signal that synchronizes a video “flash” and audio tone burst. The two signals can be monitored on an oscilloscope to determine the delay between them.

Several other types of specialized products can correct varying delays automatically, such as the Pixel AD3100 audio delay and synchronizer, which provides compensation based on a control input from a compatible video frame synchronizer. It can also automatically correct independent variable delay sources by interfacing with the company's DD2100 video delay detector, which samples the video at two points in a system and then provides a control signal to the AD3100 audio delay unit. Similarly, the Sigma Electronics Arbalest system uses a proprietary technology to provide automatic video delay detection and audio compensation in a system.

The JDSU DTS-330 real-time transport stream analyzer with SyncCheck provides lip sync analysis when used with a special video source. The K-WILL QuMax-2000 generates a “Video DNA” identifying signal that can measure the timing of video signals in a plant or even at separate locations. The Pixel Instruments LipTracker detects a face in the video and then 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 movements are analyzed to produce a measurement of the lip sync error.

VI. The Simplay Labs HD Lip Sync Performance Program

Simplay Labs is committed to enabling optimal performance of HD products, facilitating a high Quality of Experience (QoE) factor for consumers. We collaborate with retailers, content providers, and broadcast electronics professionals to enable consumers' passion for the HD lifestyle. We also work with manufacturers to ensure QoE above and beyond simple protocol specifications, by delivering the best possible HD experience to the consumer.

Evaluating the performance of a lip sync correction system requires a repeatable procedure for presenting the system with a stimulus, then quantitatively measuring its output relative to an ideal response. To meet this goal Simplay Labs has developed a complete synchronization test environment. Our custom lip sync testing techniques involve a combination of in-house tools and off-the-shelf test equipment integrated with custom software, which includes mathematics and realization programs, and specialized design packages. Sophisticated custom test fixtures were developed to implement specific test plans for Simplay HD Lip Sync performance testing.

The Simplay HD lip sync test environment system generates a custom media stimulus and presents it as input to a system under test. This custom stimulus is a Simplay Labs media sequence that requires synchronization, and must be produced in a reliable, accurate, and repeatable manner to pass our stringent requirements. In real world consumer applications, a media delivery system typically receives its input either from a player system (as in a DVD player) or from a live source (as in a broadcast application). To simulate typical consumer applications, the Simplay HD lip sync test environment is able to simulate a test sequence for sink testing.

To provide reliable measurements, a custom developed Simplay HD Lip Sync measurement media sequence is used as a controlled input, so that all timing information is known *a priori*. Therefore, given that the ideal synchronization time is known, and since the actual display time is measurable, it is possible to detect any deviations.

Simplay Labs is the only one-stop organization offering HD end-to-end lip sync performance testing services, pre-testing R&D tools, product development R&D consulting, and implementation technologies aimed at saving critical time-to-market. See *Figure 1*.

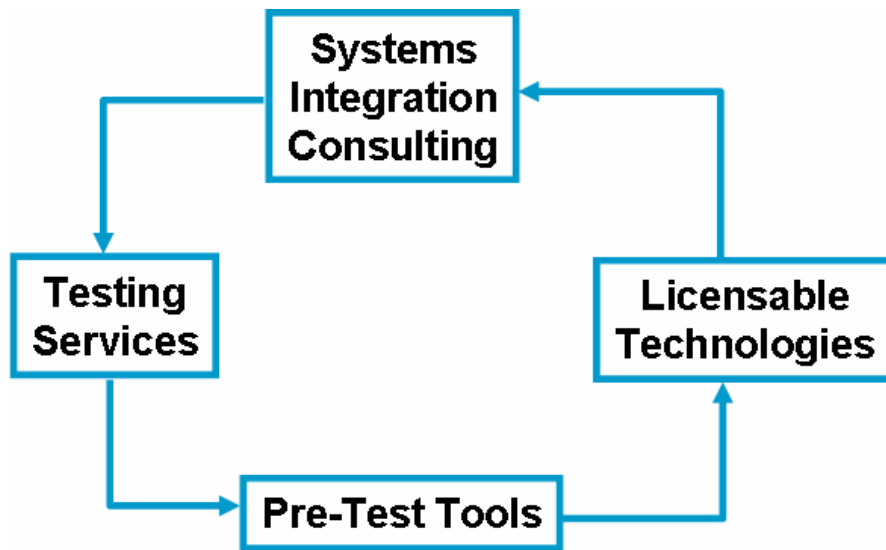


Figure 1. Simplay Labs Product & Services

All CE devices that have demonstrated adherence to the Simplay HD Lip Sync Performance Specification and passed testing by the Simplay HD Test Center are identified with the Simplay HD lip sync logo, enabling consumers to make CE equipment purchases with the confidence their HD components have been verified for peak performance.

Leveraging this branding component, the Simplay HD Testing Program is educating retail channels on the importance of lip sync performance and how to identify high QoE devices. Now and going forward, the Simplay HD mark takes the guesswork out of shopping for HD, promising the perfect performance that consumers demand. Consumers will enjoy the performance of a lifetime from their home entertainment equipment.

Copyright © 2008 Silicon Image, Inc. All rights reserved. Simplay Labs, Simplay HD and the Simplay Labs logo are trademarks or registered trademarks of Silicon Image, Inc. in the United States and/or other countries. HDMI and High-Definition Multimedia Interface are trademarks or registered trademarks of HDMI Licensing, LLC in the United States and/or other countries. All other trademarks and registered trademarks are the property of their respective owners. Product specifications are subject to change without notice.