SCTE | STANDARDS

Digital Video Subcommittee

AMERICAN NATIONAL STANDARD

ANSI/SCTE 230 2021

Recommended Practice for Proper Handling of Audio-Video Synchronization in Cable Systems

NOTICE

The Society of Cable Telecommunications Engineers (SCTE) Standards and Operational Practices (hereafter called "documents") are intended to serve the public interest by providing specifications, test methods and procedures that promote uniformity of product, interoperability, interchangeability, best practices, and the long term reliability of broadband communications facilities. These documents shall not in any way preclude any member or non-member of SCTE from manufacturing or selling products not conforming to such documents, nor shall the existence of such standards preclude their voluntary use by those other than SCTE members.

SCTE assumes no obligations or liability whatsoever to any party who may adopt the documents. Such adopting party assumes all risks associated with adoption of these documents and accepts full responsibility for any damage and/or claims arising from the adoption of such documents.

NOTE: The user's attention is called to the possibility that compliance with this document may require the use of an invention covered by patent rights. By publication of this document, no position is taken with respect to the validity of any such claim(s) or of any patent rights in connection therewith. If a patent holder has filed a statement of willingness to grant a license under these rights on reasonable and nondiscriminatory terms and conditions to applicants desiring to obtain such a license, then details may be obtained from the standards developer. SCTE shall not be responsible for identifying patents for which a license may be required or for conducting inquiries into the legal validity or scope of those patents that are brought to its attention.

Patent holders who believe that they hold patents which are essential to the implementation of this document have been requested to provide information about those patents and any related licensing terms and conditions. Any such declarations made before or after publication of this document are available on the SCTE web site at https://scte.org.

140 Philips Road
Exton, PA 19341

Document Types and Tags

Document Type: Specification

Document Tags:

☒ Test or Measurement	☒ Checklist	☐ Facility
☐ Architecture or Framework	☐ Metric	☐ Access Network
☒ Procedure, Process or Method	☐ Cloud	☐ Customer Premises

Document Release History

Release	Date
SCTE 230 2016	*3/28/2016*

Note: Standards that are released multiple times in the same year use: a, b, c, etc. to indicate normative balloted updates and/or r1, r2, r3, etc. to indicate editorial changes to a released document after the year.

Table of Contents

List of Figures

1. Introduction

1.1. Executive Summary

It has historically been a 'given' that television content is comprised of moving pictures and sound – and that the pictures and sound are presented in synchrony in the viewing / listening environment. Complex signal processing, distribution environments and consumer equipment have all conspired to make the synchronous presentation of pictures and the accompanying sound to a consumer a challenging task.

This document describes the nature of video and audio synchrony, how to recognize and measure the loss of synchrony, and potential identification and remediation steps when that synchrony is lost.

1.2. Scope

This Recommended Practice specifies proper procedures for the measurement of and maintenance of Audio-Video Synchronization (commonly known as "Lip Sync") through various aspects of a cable system – including the headend and distribution architecture and devices.

1.3. Benefits

This document assists the reader with recognition of the issue and where Lip Sync problems may have occurred and how to troubleshoot / mitigate those sync issues. Understanding the issues as outlined in this document will shorten the troubleshooting process and help to increase the ability of technical staff to communicate the issue to consumers as well as maintenance personnel.

1.4. Intended Audience

This document is intended for technical operations engineering and, potentially, customer contact personnel such as installers to be able to identify audio/video synchronization problems.

2. Normative References

The following documents contain provisions which, through reference in this text, constitute provisions of this document. The editions indicated were valid at the time of subcommittee approval. All documents are subject to revision and, while parties to any agreement based on this document are encouraged to investigate the possibility of applying the most recent editions of the documents listed below, they are reminded that newer editions of those documents might not be compatible with the referenced version.

2.1. SCTE References

[SCTE 197] ANSI/SCTE 197 2018, "Recommendations for Spot Check Loudness Measurements,"

2.2. Standards from Other Organizations

[CEB-20] CTA CEB-20 R-2013, "A/V Synchronization Processing Recommended Practice,"

[BT1359] ITU BT.1359, "Relative timing of sound and vision for broadcasting,"

[ST2064-1] SMPTE ST2064-1:2015, "Audio to Video Synchronization Measurement – Fingerprint Generation," https://www.smpte.org/standards/document-index

[ST2064-2] SMPTE ST2064-2:2015, "Audio to Video Synchronization Measurement – Fingerprint Transport," https://www.smpte.org/standards/document-index

2.3. Other Published Materials

No normative references are applicable.

3. Informative References

The following documents might provide valuable information to the reader but are not required when complying with this document.

3.1. SCTE References

No informative references are applicable.

3.2. Standards from Other Organizations

No informative references are applicable.

3.3. Other Published Materials

No informative references are applicable.

4. Compliance Notation

shall	This word or the adjective "***required***" means that the item is an absolute requirement of this document.
shall not	This phrase means that the item is an absolute prohibition of this document.
forbidden	This word means the value specified ***shall*** never be used.
should	This word or the adjective "*recommended*" means that there *may* exist valid reasons in particular circumstances to ignore this item, but the full implications *should* be understood and the case carefully weighed before choosing a different course.
should not	This phrase means that there *may* exist valid reasons in particular circumstances when the listed behavior is acceptable or even useful, but the full implications *should* be understood and the case carefully weighed before implementing any behavior described with this label.
may	This word or the adjective "*optional*" indicate a course of action permissible within the limits of the document.
deprecated	Use is permissible for legacy purposes only. Deprecated features *may* be removed from future versions of this document. Implementations *should* avoid use of deprecated features.

5. Abbreviations and Definitions

5.1. Abbreviations

AC-3	Audio Codec 3
CTA	Consumer Technology Association
Lip Sync	Audio-Video Synchronization
SCTE	Society of Cable Telecommunications Engineers
SMPTE	Society of Motion Picture and Television Engineers

5.2. Definitions

Definitions of terms used in this document are provided in this section. Defined terms that have specific meanings are capitalized. When the capitalized term is used in this document, the term has the specific meaning as defined in this section.

downstream	The direction of signal transmission from the headend or hub site to the subscriber. Also called forward.
upstream	The direction of signal transmission from the subscriber to the hub site or headend. Also called return or reverse.

6. Audio Video Synchronization Background and Detection

6.1. Detection and Visibility

Today's complex 'source to sink' flow of digital audio and video includes many active processing steps that *may* result in undesired changes to the relative timing of video to the related audio. This relationship is commonly called "Lip Sync" – and it is annoying to viewers when the video and audio elements of a program are not synchronized within some small tolerable time difference. Lip Sync errors become apparent when a speaker's lips are visibly moving while their voice is heard. Maintaining proper Lip Sync has been a challenge since the earliest motion pictures with sound were produced (including phonograph records accompanying a film evolving through sound-on-film projector loops to a separate sound head).

Concerns about Lip Sync errors have remained through analog television to the digital transition and codecs and transport systems.

In fact, the 'transition' to digital displays has made Lip Sync errors potentially worse. The end-to-end delivery path for analog television, including the viewer's CRT-based television set, had a well-known deterministic latency for the video and audio paths to enable minimization of lip sync errors. Contemporary fixed-pixel displays present a much greater degree of uncertainty in relative latency between audio and video as they almost always contain frame-buffers, frame-rate conversion, temporal noise reduction and other signal processing that can introduce significant Latency, including variations in the Latency that are not consistent nor known to external audio equipment. Existing standards (ITU BT.1359) [BT1359] were made with standard definition CRTs not HD fixed-pixel displays. No update has been done to the existing ITU standard since the introduction of these new displays.

Television program distributors want to minimize Lip Sync errors for their customers and so require the means to detect and measure the relative latency between audio and video program elements. Customer complaints often include subjective observations that "something doesn't look right". Investigating such complaints includes a determination of which factors are within the control of the distributor and should be addressed appropriately.

6.1.1. Objective Measurements

Objective synchrony measurements are possible with test patterns and audio tones as an out-of-service process – which may not be via a production workflow or signal path, potentially bypassing the malfunctioning path.

Examples of objective measurements are the VALID test signal (as-of this writing, used by HBO during monthly transmission tests; see: homeboxoffice.com (TechOps section), for 2016: https://www.homeboxoffice.com/TechOps%20Docs/2016_Transmission_Tests%20w_Logos.pdf)

General information about the VALID device: http://www.visuals-switzerland.net/attachment.php?id_attachment=2487

6.1.2. Visual / Aural Observation

Trained observers can detect Lip Sync errors within a few frames. This document does not attempt to provide a full course in how to observe errors to such a precise level, but certain visual cues in program content are easy to see and can be used to trigger further investigation of a potential problem. Such observations could include: Visual cues such as a gun-shot (sound and picture *should* be matched) or drum thump / symbol crash with drumstick visible are good clues that *may* conveniently occur during a brief period of watching a program. Generally, audio dialog is more subtle and can be impacted by dubbing and dialog replacement during production processes. Also note that some production choices include items that *may* be improperly identified as an error such as a distant explosion that would include a delay – as sound is normally delayed in physics / nature.

Specific test materials can be authored (e.g. a VOD asset) and made available to consumers to have a "reference" known asset to visually verify (for non-expert viewers) that Lip Sync is correct and other troubleshooting is warranted.

6.2. Synchronization Error Budget

Several human-factors studies have been done to determine visibility and acceptability for Lip Sync errors. A "pass-fail" specification is beyond the scope of this document. However, the threshold of "consumer complaints" is generally a threshold that *should* not be exceeded. Thus, the following example from ITU-R BT.1359 [BT1359] *may* be used for guidance, noting the caution at the beginning of this section concerning the impact of fixed-pixel displays on this recommended practice.

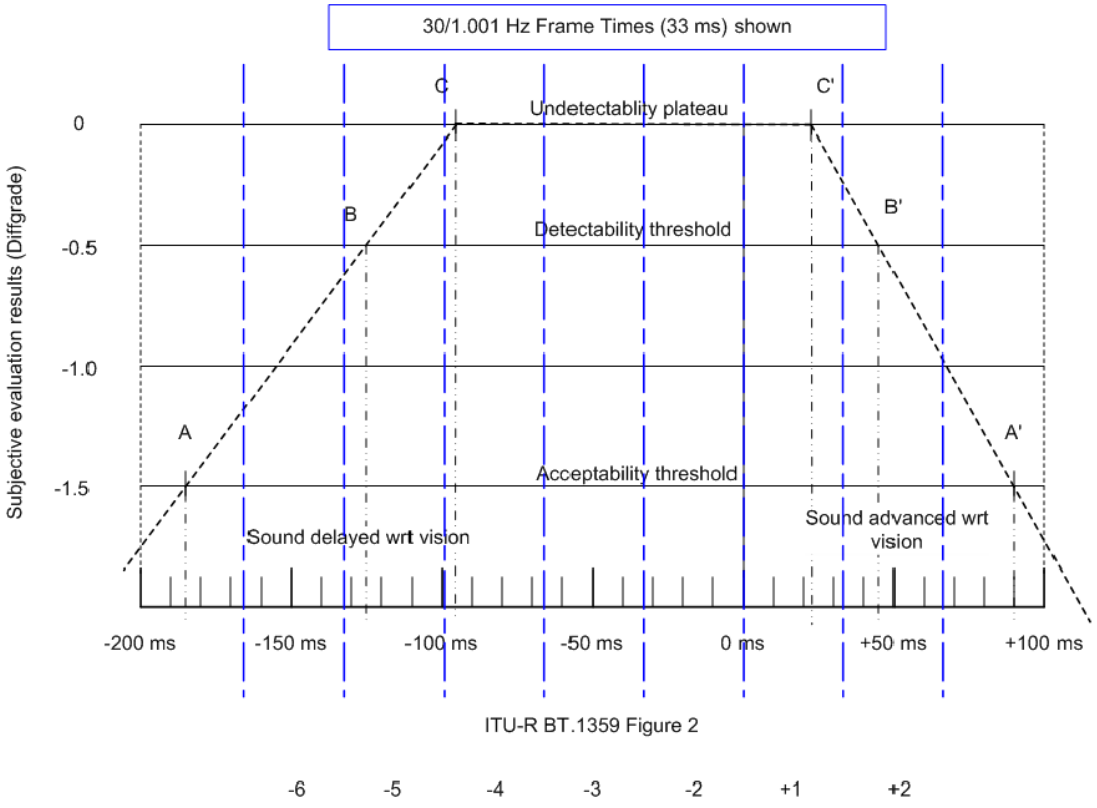

Figure 1 - Lip Sync Visibility and Acceptability

7. Audio Video Synchronization Signal Flow and Processing

7.1. Introduction

This Recommended Practice outlines the flow of audio-video content through a digital cable system and what procedures can be applied to properly maintain Lip Sync along the 'journey' to the ultimate viewing of the audio and video.

7.2. Overview

Figure 2 depicts a simplified flow of the various forms of digital audio and video that *may* be acquired by a cable operator for processing and distribution. Section "A" of the drawing represents "Content Acquisition" – where the operator "acquires" the streams and *may* process both video and audio. Section "B" lists some, but not all, of the processing functions that *may* occur in a "Headend" (other terms might include Network Operations Center (NOC) or other facility) – each one of these processing steps *may* have an impact on Lip Sync. Section "C" reflects the concept of distribution – which *may* or may not be under the operator's control and *may* distribute content to devices that *may* or may not be under the complete control of the operator.

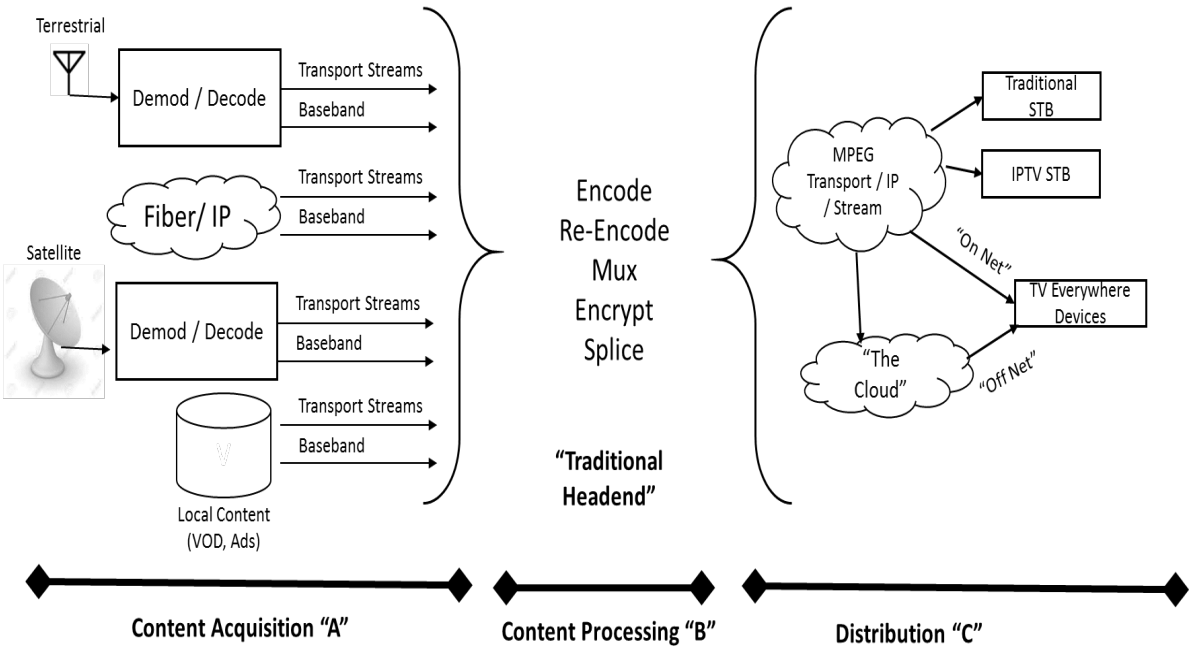

Figure 2 - Signal Flow Overview

7.3. Content Acquisition

As shown in Figure 2, in simplified summary, there are numerous sources and formats of audio and video content. Ideally, all sources of content *should* be received with the proper technical specifications (e.g. MPEG timing) and, ideally, the content *should* have proper Lip Sync.

Content Lip Sync can be verified by various techniques such as out-of-service test patterns and tones, and in some cases, observation by an expert viewing (with the presumption that content was created correctly and not dubbed or contain improper dialog replacement, etc.)

If Lip Sync errors are detected at the acquisition stage, the operator *should* contact the program provider and work with the provider and equipment vendor(s) to resolve technical problems resulting in Lip Sync errors.

Content *may* include techniques such as SMPTE ST 2064 [ST2064-1, ST2064-2] Lip Sync fingerprinting which provides in-service measurement of sync errors. If SMPTE fingerprinting data is present, the operator *may* use equipment or functionality in processing steps to correct for errors, if available.

7.4. Content processing

During content processing, operators and vendors *should* use due care to ensure that Lip-Sync is maintained during the various functions of the distribution process. Numerous technical specifications apply to the functions that are performed in typical headend facilities as defined by SMPTE, ATSC, SCTE and CTA. For the purposes of this document, these various specifications *may* be implemented with a concern toward proper Lip Sync. It is broader than the scope of this document to enumerate the specific application of these specifications to individual signal processing functions.

7.5. Distribution

Operator and consumer-owned devices *should* follow good practice to minimize Lip Sync errors during distribution, decoding, rendering and display. Sink devices *should* adhere to CTA CEB-20 [CEB-20].

It is important to note that consumer in-home equipment can be quite complex. A 'simple' configuration of a STB connected to a TV display via HDMI *may* be the most straightforward use case (and used millions of times) *may* still have Lip Sync errors. Home theater installations introduce additional complexity. It *may* be necessary to use a 'simple case' for troubleshooting.

Also note that TV displays and home theater devices might have "delay adjustments" that permanently change Lip Sync. Those adjustments could be initially miss-set. System-wide delay settings *should not* be used to correct single-channel or single-program errors.

8. Corrective Measures

8.1. Determination of Magnitude

Generally, errors *may* be insertion-related (i.e. ad content that *may* be locally-inserted), program related (i.e. a single event on a network) or continual over a program feed. Noting the duration and severity of events is useful in troubleshooting and reporting problems to program suppliers or vendors.

If a Lip Sync error is encountered that warrants escalation, the signal path and "likely suspect" functional process *should* be noted and used to begin a troubleshooting next-step(s).

8.2. Vendor and Signal Processing

A likely approach to troubleshooting would be to compare a signal path (i.e. multiplexer, encoding thread, etc.) assigned to one network to an identical signal path (including identical equipment configuration) assigned to a different network. Identical equipment *may* exhibit similar errors (unless Lip Sync errors are due to accumulated drift of clocks or fixed by fresh acquisition of the program service (e.g. tune-away and back)). This troubleshooting could result in a bug-report to a vendor.

8.3. Program Supplier

On egregious single-program errors or long-term feed errors that cannot be identified by equipment issues, a viable next-step would be to call the program provider and supply the detailed observation and troubleshooting steps taken. The SCTE 197 (Section 8) [[SCTE 197]] also discusses development of a criterion for contacting a program provider / supplier with similar audio-related issues.

8.4. Documentation

Applied generally to the items in this Section 7, it is important to document the content source, program or asset name, time and date, equipment and configuration used (Section 7.1). Additionally, a recording of specific examples *may* be useful (Sections 7.2 and 7.3) to use with vendors or program suppliers. "Sample Points" for such recordings *may* follow Figure 2, Sections "A", "B", or "C".

Annex A. Example Cases of Sync Errors

This informative Annex provides examples of a typical audio/video signal flow for both a "Reference Case" (no sync errors) and a "Pathological Example" of typical locations in a signal path where errors can occur.

It is important to note that sync errors can occur in several locations (and in different lead/lag combinations) that *may* be transient in nature. It is beyond the scope of this document to provide guidance for every scenario that *may* result in the end-consumer observing noticeable Sync errors.

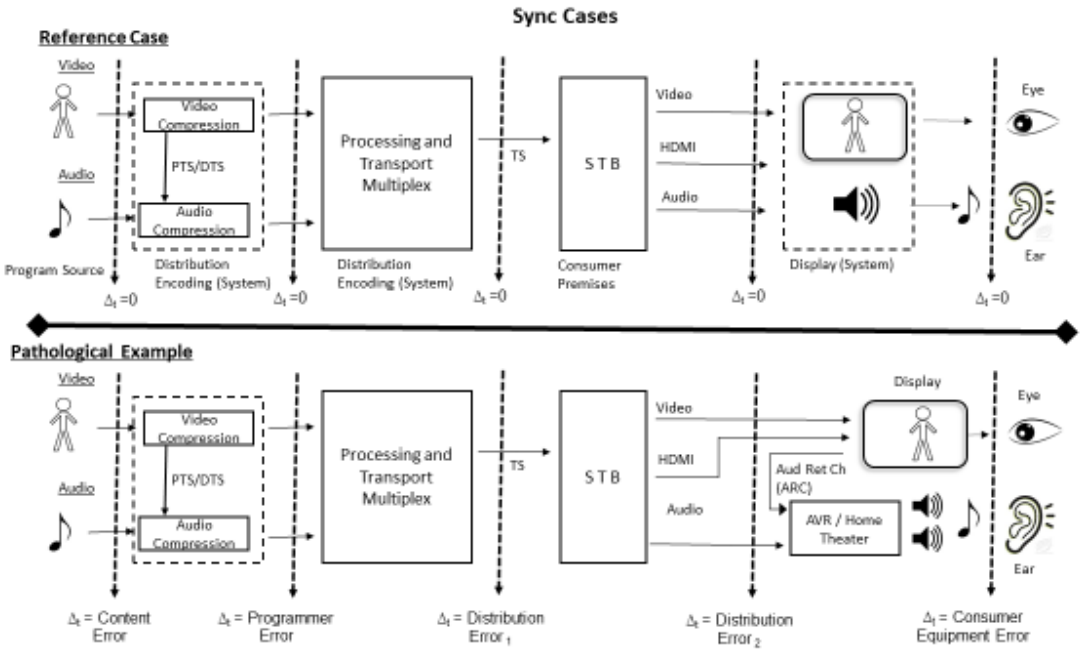

Figure 3 - Sync Cases

The Figure *3* Sync Cases provides highly simplistic notions of how Lip Sync errors can occur and where those errors can impact downstream systems and, ultimately, the consumer experience.

A narrative example of those errors in Figure 3 might include these or other descriptions:

Δ_t = Content Error: The creative process itself *may* deliver content that may not be or *appears* 'not-in-sync' – this could be a result of dialog replacement or other subtle (or not so subtle) sync errors. Some content "creation" such as remote broadcasts during local news coverage *may* be out-of-sync and cannot easily be corrected during the production process.

Δ_t = Programmer Error: Failure during the playback, switching and integration of content *may* result in Sync errors either on short (commercials) or long-form (program) content.

Δ_t = Distribution Error (1): At the national, regional, and local levels there are many steps in the processing of program streams and advertising inserts therein. Multiplexing, demultiplexing and transport stream manipulation can cause inadvertent loss of sync.

Δ_t = Distribution Error (2): In the consumer premises the termination of the cable network i.e. the "delivery point" is either baseband video, analog or digital audio in several formats (e.g. S/P-DIF, TOSLINK, HDMI).This diagram omits additional complexity such as TV Everywhere and/or consumer device "apps" and 'whole-home' scenarios. The STB *may* introduce sync errors, perhaps based on differential delays among the various interfaces and not knowing what device(s) follow the STB delivery.

Δ_t = Consumer Equipment Error: The ultimate complexity exists in the equipment configuration that the cable operator has little to no control over. The near-infinite connection and configuration options result in a troubleshooting nightmare. Differential Sync delays among decoding modes, speaker placement, display-to-AVR (Audio Video Receiver) "Audio Return Channel" (ARC) configuration, etc. are beyond the scope of description here. Consumer Equipment Error is the most difficult to diagnose as even experienced field personnel cannot be expected to know every complex brand, model, and consumer setup variation.