SMPTE STANDARD

Audio to Video Synchronization Measurement — Fingerprint Generation

Table of Contents

Approved
October 9, 2015

Foreword

The Society of Motion Picture and Television Engineers (SMPTE) is an internationally-recognized standards developing organization. Headquartered and incorporated in the United States of America, SMPTE has members in over 80 countries on six continents. SMPTE's Engineering Documents, including Standards, Recommended Practices, and Engineering Guidelines, are prepared by SMPTE's Technology Committees. Participation in these Committees is open to all with a bona fide interest in their work. SMPTE cooperates closely with other standards-developing organizations, including ISO, IEC and ITU.

SMPTE Engineering Documents are drafted in accordance with the rules given in the Standards Operations Manual.

SMPTE ST 2064-1 was prepared by Technology Committee 24TB.

Intellectual Property

SMPTE draws attention to the fact that it is claimed that compliance with this Standard may involve the use of one or more patents or other intellectual property rights (collectively, "IPR"). The Society takes no position concerning the evidence, validity, or scope of this IPR.

Each holder of claimed IPR has assured the Society that it is willing to License all IPR it owns, and any third party IPR it has the right to sublicense, that is essential to the implementation of this Standard to those (Members and non-Members alike) desiring to implement this Standard under reasonable terms and conditions, demonstrably free of discrimination. Each holder of claimed IPR has filed a statement to such effect with SMPTE. Information may be obtained from the Director, Standards & Engineering at SMPTE Headquarters.

Attention is also drawn to the possibility that elements of this Standard may be subject to IPR other than those identified above. The Society shall not be responsible for identifying any or all such IPR.

Introduction

This section is entirely informative and does not form an integral part of this Engineering Document.

Errors in audio to video timing relationships have become commonplace in the industry. Such errors are known as "lip-sync" errors because of a typical viewer's sensitivity to inaccuracy in the synchronization of lip movement with the sound of speech. This situation arises because the audio and video portions of a program follow different processing paths, each with its own inherent timing factors. For various reasons, the audio and video delays through the program distribution chain may change dynamically. In order to correct such timing errors as they occur, a method is required to measure the change in audio to video synchronization without relying on out-of-service test signals.

The SMPTE 2064 suite of documents for Audio to Video Synchronization Measurements defines a process for extracting, packetizing, and transporting compact representations of audio and video essence, known as video and audio fingerprints, which change constantly as a function of changing picture and sound content. These fingerprints are extracted by non-intrusive analysis of the audio and video essence, and therefore can be used in a live on-air situation as well as in non-real-time systems. The fingerprints generated at a reference point where synchronization is known to be correct are intended to be transported through the program distribution chain, either bound to or separate from the associated essence. Where measurement and correction of synchronization errors is needed, a new set of fingerprints is extracted from the essence at

a downstream location and compared with the fingerprints from the reference point. This comparison provides a dynamic measurement of the audio to video timing changes that have occurred, which may be used by other processes to display and/or correct any synchronization errors.

The timing relationship of the audio and video signals within a program is described from the perspective of the video signal. When the audio signal precedes, or arrives earlier in time than the video signal the measured value is stated as a negative value. When the audio signal arrives later in time than the video signal it is stated as a positive value. This measurement and comparison is outside the scope of this document.

The SMPTE ST 2064-1 standard specifies the method for generating the audio and video fingerprints, and the SMPTE ST 2064-2 standard specifies the carriage of fingerprints using various transport methods.

1 Scope

This standard defines algorithms and procedures for generating audio and video fingerprints from audio and video essence used for audio to video timing measurements. It also specifies a method for combining the audio and video fingerprints and associated metadata into a container suitable for transport. Composite video formats are not supported by this standard.

2 Conformance Notation

Normative text is text that describes elements of the design that are indispensable or contains the conformance language keywords: "shall", "should", or "may". Informative text is text that is potentially helpful to the user, but not indispensable, and can be removed, changed, or added editorially without affecting interoperability. Informative text does not contain any conformance keywords.

All text in this document is, by default, normative, except: the Introduction, any section explicitly labeled as "Informative" or individual paragraphs that start with "Note:"

The keywords "shall" and "shall not" indicate requirements strictly to be followed in order to conform to the document and from which no deviation is permitted.

The keywords, "should" and "should not" indicate that, among several possibilities, one is recommended as particularly suitable, without mentioning or excluding others; or that a certain course of action is preferred but not necessarily required; or that (in the negative form) a certain possibility or course of action is deprecated but not prohibited.

The keywords "may" and "need not" indicate courses of action permissible within the limits of the document.

The keyword “reserved” indicates a provision that is not defined at this time, shall not be used, and may be defined in the future. The keyword “forbidden” indicates “reserved” and in addition indicates that the provision will never be defined in the future.

A conformant implementation according to this document is one that includes all mandatory provisions ("shall") and, if implemented, all recommended provisions ("should") as described. A conformant implementation need not implement optional provisions ("may") and need not implement them as described.

Unless otherwise specified, the order of precedence of the types of normative information in this document shall be as follows: Normative prose shall be the authoritative definition; Tables shall be next; followed by formal languages; then figures; and then any other language forms.

3 Normative References

Note: All references in this document to other SMPTE documents use the current numbering style (e.g. SMPTE ST 274:2008) although, during a transitional phase, the document as published (printed or PDF) may bear an older designation (such as SMPTE 274M-2008). Documents with the same root number (e.g. 274) and publication year (e.g. 2008) are functionally identical.

The following standards contain provisions which, through reference in this text, constitute provisions of this Standard. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this Standard are encouraged to investigate the possibility of applying the most recent edition of the standards indicated below.

SMPTE ST 125:2013, SDTV Component Video Signal Coding 4:4:4 and 4:2:2 for 13.5 MHz and 18 MHz Systems