
**Information technology — Coding of
audio-visual objects —**

**Part 15:
Carriage of network abstraction layer
(NAL) unit structured video in the ISO
base media file format**

*Technologies de l'information — Codage des objets audiovisuels —
Partie 15: Transport de vidéo structuré en unités NAL au format ISO
de base pour les fichiers médias*

ISO/IEC 14496-15:2017

<https://standards.iteh.ai/catalog/standards/iso/263729b6-10f0-49a5-bd17-421f87ed0568/iso-iec-14496-15-2017>



iTeh Standards
(<https://standards.iteh.ai>)
Document Preview

ISO/IEC 14496-15:2017

<https://standards.iteh.ai/catalog/standards/iso/263729b6-10f0-49a5-bd17-421f87ed0568/iso-iec-14496-15-2017>



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2017, Published in Switzerland

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

Contents

Page

Foreword	vi
Introduction	vii
1 Scope	1
2 Normative references	1
3 Terms, definitions and abbreviated terms	1
3.1 Terms and definitions	1
3.2 Abbreviated terms	7
4 General definitions	8
4.1 Overview	8
4.2 Elementary stream structure	8
4.3 Sample and configuration definition	8
4.3.1 General	8
4.3.2 Canonical order and restrictions	9
4.3.3 Sample format	9
4.3.4 Optional boxes in the sample entry	10
4.4 Video track structure	10
4.5 Template fields used	11
4.6 Visual width and height	11
4.7 Decoding time (DTS) and composition time (CTS)	11
4.8 Sample groups on random access recovery points 'roll' and random access points 'rap'	11
4.9 Hinting	12
4.10 On change of sample entry	12
5 AVC elementary streams and sample definitions	13
5.1 General	13
5.2 Elementary stream structure	14
5.3 Sample and configuration definition	16
5.3.1 General	16
5.3.2 Canonical order and restrictions	16
5.3.3 Decoder configuration information	17
5.4 Derivation from ISO base media file format	20
5.4.1 AVC file type and identification	20
5.4.2 AVC video stream definition	20
5.4.3 AVC parameter set stream definition	21
5.4.4 Parameter sets	22
5.4.5 Sync sample	23
5.4.6 Shadow sync	23
5.4.7 Layering and sub-sequences	23
5.4.8 Alternate streams and switching pictures	27
5.4.9 Definition of a sub-sample for AVC	29
5.4.10 Post-decoder requirements scheme for signalling of SEI for AVC	29
6 SVC elementary stream and sample definitions	30
6.1 General	30
6.2 Elementary stream structure	31
6.3 Use of the plain AVC file format	32
6.4 Sample and configuration definition	32
6.4.1 General	32
6.4.2 Canonical order and restrictions	32
6.5 Derivation from the ISO base media file format	34
6.5.1 SVC track structure	34
6.5.2 Data sharing and extraction	34
6.5.3 SVC video stream definition	35

6.5.4	SVC visual width and height.....	37
6.5.5	Sync sample.....	37
6.5.6	Shadow sync.....	37
6.5.7	Independent and disposable samples box.....	37
6.5.8	Sample groups on random access recovery points 'roll' and random access points 'rap'.....	38
6.5.9	Definition of a sub-sample for SVC.....	38
7	MVC and MVD elementary stream and sample definitions.....	39
7.1	General.....	39
7.2	Overview of MVC or MVD Storage.....	40
7.3	MVC and MVD elementary stream structures.....	42
7.4	Use of the plain AVC file format.....	43
7.5	Sample and configuration definition.....	44
7.5.1	General.....	44
7.5.2	Canonical order and restriction.....	44
7.5.3	Decoder configuration record.....	44
7.6	Derivation from the ISO base media file format.....	46
7.6.1	MVC and MVD track structures.....	46
7.6.2	Reconstruction of an access unit.....	47
7.6.3	Sample entry.....	48
7.6.4	Sync sample.....	58
7.6.5	Shadow sync.....	59
7.6.6	Independent and disposable samples box.....	59
7.6.7	Sample groups on random access recovery points 'roll' and random access points 'rap'.....	59
7.7	MVC specific information boxes.....	59
7.7.1	General.....	59
7.7.2	Multiview information box.....	60
7.7.3	Multiview group box.....	60
7.7.4	Multiview group relation box.....	62
7.7.5	Multiview relation attribute box.....	62
7.7.6	Multiview scene info box.....	67
7.7.7	MVC view priority assignment box.....	68
8	HEVC elementary streams and sample definitions.....	68
8.1	General.....	68
8.2	Elementary stream structure.....	69
8.3	Sample and configuration definition.....	69
8.3.1	General.....	69
8.3.2	Canonical order and restrictions.....	69
8.3.3	Decoder configuration information.....	70
8.4	Derivation from ISO base media file format.....	73
8.4.1	HEVC video stream definition.....	73
8.4.2	Parameter sets in sample entry.....	74
8.4.3	Sync sample.....	74
8.4.4	Sync sample sample grouping.....	75
8.4.5	Temporal scalability sample grouping.....	75
8.4.6	Temporal sub-layer access sample grouping.....	77
8.4.7	Step-wise temporal layer access sample grouping.....	77
8.4.8	Definition of a sub-sample for HEVC.....	78
8.4.9	Handling non-output samples.....	80
9	Layered HEVC elementary stream and sample definitions.....	80
9.1	General.....	80
9.2	Overview of L-HEVC storage.....	81
9.3	L-HEVC elementary stream structure.....	82
9.4	Sample and configuration definition.....	82
9.4.1	General.....	82
9.4.2	Canonical order and restrictions.....	82

9.4.3	Decoder configuration record	82
9.5	Derivation from the ISO base media file format and the HEVC file format (Clause 8)	83
9.5.1	L-HEVC track structure	83
9.5.2	Data sharing and reconstruction of an L-HEVC bitstream	84
9.5.3	L-HEVC video stream definition	85
9.5.4	L-HEVC visual width and height	88
9.5.5	Sync sample	88
9.5.6	Independent and disposable samples box	89
9.5.7	Stream access point sample group	89
9.5.8	The 'roll', 'rap', 'sync', 'tsas' and 'stsa' sample groups	89
9.5.9	Definition of a sub-sample for L-HEVC	90
9.5.10	Handling non-output samples	90
9.6	L-HEVC specific structures	90
9.6.1	External base layer sample group	90
9.6.2	The operating points information sample group	91
9.6.3	The layer information sample group	94
10	Storage of tiled HEVC and L-HEVC video streams	95
10.1	General	95
10.2	NAL unit map entry	95
10.2.1	Definition	95
10.2.2	Syntax	96
10.2.3	Semantics	96
10.3	Tile region group entry	97
10.3.1	Definition	97
10.3.2	Syntax	97
10.3.3	Semantics	97
10.4	Tile sub track definition	99
10.4.1	General	99
10.4.2	TileSubTrackGroupBox	99
10.5	HEVC and L-HEVC tile track	99
10.5.1	General	99
10.5.2	Sample entry name and format for HEVC tile tracks	100
10.5.3	Sample entry name and format for L-HEVC tile tracks	101
10.5.4	Bitstream reconstruction from tile base and tile tracks	102
10.5.5	Sample entry names for tile base tracks	102
	Annex A (normative) In-stream structures	104
	Annex B (normative) SVC, MVC, and MVD sample group and sub-track definitions	113
	Annex C (normative) Temporal metadata support	134
	Annex D (normative) File format toolsets and brands	143
	Annex E (normative) Sub-parameters for the MIME type “codecs” parameter	145

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see the following URL: www.iso.org/iso/foreword.html.

The committee responsible for this document is ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

This fourth edition cancels and replaces the third edition (ISO/IEC 14496-15:2014), which has been technically revised.

It also incorporates the Technical Corrigendum ISO/IEC 14496-15:2014/Cor 1:2015.

A list of all parts in the ISO/IEC 14496 series can be found on the ISO website.

Introduction

This document defines a storage format based on, and compatible with, the ISO Base Media File Format (ISO/IEC 14496-12), which is used by the MP4 file format (ISO/IEC 14496-14) and the Motion JPEG 2000 file format (ISO/IEC 15444-3) among others. This document enables video streams formatted as Network Adaptation Layer Units (NAL Units) to

- a) be used in conjunction with other media streams, such as audio,
- b) be used in an MPEG-4 systems environment, if desired,
- c) be formatted for delivery by a streaming server, using hint tracks, and
- d) inherit all the use cases and features of the ISO Base Media File Format on which MP4 and MJ2 are based.

This document may be used as a standalone specification; it specifies how NAL unit structured video content shall be stored in an ISO Base Media File Format compliant format. However, it is normally used in the context of a specification, such as the MP4 file format, derived from the ISO Base Media File Format, that permits the use of NAL unit structured video such as AVC (ISO/IEC 14496-10) video and High Efficiency Video Coding (HEVC, ISO/IEC 23008-2) video.

The ISO Base Media File Format is becoming increasingly common as a general-purpose media container format for the exchange of digital media, and its use in this context should accelerate both adoption and interoperability.

The International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC) draw attention to the fact that it is claimed that compliance with this document may involve the use of a patent.

The ISO and IEC take no position concerning the evidence, validity and scope of this patent right.

The holder of this patent right has assured the ISO and IEC that he is willing to negotiate licences under reasonable and non-discriminatory terms and conditions with applicants throughout the world. In this respect, the statement of the holder of this patent right is registered with the ISO and IEC.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Information technology — Coding of audio-visual objects —

Part 15: Carriage of network abstraction layer (NAL) unit structured video in the ISO base media file format

1 Scope

This document specifies the storage format for streams of video that is structured as NAL units, such as AVC (ISO/IEC 14496-10) and HEVC (ISO/IEC 23008-2) video streams.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 14496-12:2015, *Information technology — Coding of audio-visual objects — Part 12: ISO base media file format*

ISO/IEC 14496-10:2014, *Information technology — Coding of audio-visual objects — Part 10: Advanced Video Coding*

ISO/IEC 23008-2:—¹⁾, *Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 2: High efficiency video coding*

3 Terms, definitions and abbreviated terms

3.1 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 14496-10 or ISO/IEC 23008-2, and the following apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

- IEC Electropedia: available at <http://www.electropedia.org/>
- ISO Online browsing platform: available at <http://www.iso.org/obp>

3.1.1

3D-AVC NAL unit

3D-AVC VCL NAL unit

NAL unit with type 21 with `avc_3d_extension_flag` equal to 1

Note 1 to entry: As specified in ISO/IEC 14496-10:2014, Annex J.

1) To be published.

3.1.2

aggregator

in-stream structure (3.1.10) using a NAL unit header for grouping of NAL units belonging to the same sample

3.1.3

AVC base layer

maximum subset of a bitstream that is AVC compatible (i.e. a bitstream not using any of the functionality of ISO/IEC 14496-10:2014, Annex G, Annex H, Annex I, or Annex J)

Note 1 to entry: The AVC base layer is represented by AVC VCL NAL units and associated non-VCL NAL units.

Note 2 to entry: The AVC base layer itself can be a temporal scalable bitstream.

3.1.4

AVC NAL unit

AVC VCL NAL unit (3.1.5) and its associated non-VCL NAL units in a bitstream

3.1.5

AVC VCL NAL unit

NAL unit with type 1 to 5 (inclusive)

Note 1 to entry: As specified in ISO/IEC 14496-10.

3.1.6

complete subset

minimal set of tracks that contain all the information in the original bitstream

3.1.7

cropped frame dimensions

width and height of the decoded frame after applying the output cropping parameters specified by the active SPS

3.1.8

extraction path

set of operations on the original bitstream, each yielding a subset bitstream, ordered such that the complete bitstream is first in the set, and the base layer is last, and all the bitstreams are in decreasing complexity (along one of the scalability axes, such as resolution), and where every bitstream is a valid operating point

Note 1 to entry: An extraction path may be represented by the values of *priority_id* in the NAL unit headers. Alternatively, an extraction path can be represented by the run of tiers or by a set of hierarchically dependent tracks.

3.1.9

extractor

in-stream structure (3.1.10) using a NAL unit header for extraction of data from other tracks

Note 1 to entry: Extractors contain instructions on how to extract data from other tracks. Logically, an extractor can be seen as a pointer to data. While reading a track containing extractors, the extractor is replaced by the data it is pointing to.

3.1.10

in-stream structure

structure residing within sample data

3.1.11

layer set

set of layers represented within a bitstream created from another bitstream by operation of the sub-bitstream extraction process

Note 1 to entry: As specified in ISO/IEC 23008-2.

3.1.12**MVC NAL unit**

MVC VCL NAL unit (3.1.13) and its associated non-VCL NAL units in an MVC stream

Note 1 to entry: As specified in ISO/IEC 14496-10:2014, Annex H.

3.1.13**MVC VCL NAL unit**

NAL unit with type 20, and NAL units with type 14 when the immediately following NAL units are AVC VCL NAL units

Note 1 to entry: As specified in ISO/IEC 14496-10.

Note 2 to entry: MVC VCL NAL units do not affect the decoding process of a legacy AVC decoder.

3.1.14**MVC+D depth NAL unit****MVC+D depth VCL NAL unit**

NAL unit with type 21 containing a coded slice extension for a depth view component

Note 1 to entry: As specified in ISO/IEC 14496-10:2014, Annex I.

3.1.15**MVD NAL unit****MVD VCL NAL unit**

NAL unit with type 21, containing a coded slice extension for a depth view component coded with MVC+D or 3D-AVC, or a 3D-AVC texture view component

Note 1 to entry: As specified in ISO/IEC 14496-10:2014, Annex I or Annex J.

3.1.16**NAL-unit-like structure**

data structure that is similar to NAL units in the sense that it also has a NAL unit header and a payload, with a difference that the payload might not follow the start code emulation prevention mechanism required for the NAL unit syntax

Note 1 to entry: As specified in ISO/IEC 14496-10 or ISO/IEC 23008-2.

3.1.17**natively present**

not included in an *aggregator* (3.1.2) or an *extractor* (3.1.9)

Note 1 to entry: Data referred to by (hence not included in) an aggregator is considered as natively present. Data included in an aggregator is not considered as natively present.

3.1.18**operating point**

independently decodable subset of a layered bitstream

Note 1 to entry: For a layered HEVC elementary stream or tiled HEVC and L-HEVC video streams, one or more layers in the set of layers are indicated to be output layers.

Note 2 to entry: Each operating point consists of all the data needed to decode this particular bitstream subset.

Note 3 to entry: In an SVC stream, an operating point represents a particular spatial resolution, temporal resolution, and quality, and can be represented either by (i) specific values of DTQ (dependency_id, temporal_id and quality_id) or (ii) specific values of P (priority_id) or (iii) combinations of them (e.g. PDTQ). Note that the usage of priority_id is defined by the application. In an SVC file, a track represents one or more operating points. Within a track, tiers may be used to define multiple operating points.

Note 4 to entry: The bitstream subset of an MVC or MVD operating point represents a particular set of target output views at a particular temporal resolution, and consists of all the data needed to decode this particular bitstream subset. In MVD, each target output view in the bitstream subset of an MVD operating point may contain a texture view, a depth view or both.

Note 5 to entry: An operating point is referred to as an operation point in ISO/IEC 14496-10:2014, Annex H or an output operation point in ISO/IEC 23008-2.

3.1.19

output layer set

set of layers consisting of the layers of one of the specified *layer sets* (3.1.11), where one or more layers in the set of layers are indicated to be output layers

Note 1 to entry: As specified in ISO/IEC 23008-2.

3.1.20

parameter set

video parameter set, sequence parameter set, or picture parameter set, as defined in the applicable video standard (e.g. ISO/IEC 14496-10 or ISO/IEC 23008-2)

Note 1 to entry: This term is used to refer to all types of parameter sets.

3.1.21

parameter set elementary stream

elementary stream containing samples made up of only sequence and picture parameter set NAL units synchronized with the *video elementary stream* (3.1.38)

3.1.22

picture unit

set of VCL NAL units and their associated non-VCL NAL units

Note 1 to entry: As specified in ISO/IEC 23008-2.

3.1.23

prefix NAL unit

NAL units with type 14

Note 1 to entry: As specified in ISO/IEC 14496-10.

Note 2 to entry: Prefix NAL units provide scalability information about AVC VCL NAL units and filler data NAL units. Prefix NAL units do not affect the decoding process of a legacy AVC decoder. The behaviour of a legacy AVC file reader as a response to prefix NAL units is undefined.

3.1.24

reference layer

layer that is indicated as possibly needed for decoding of another layer

Note 1 to entry: As specified in ISO/IEC 23008-2 and as specified by the 'oinf' sample group defined in 9.6.2.

3.1.25

scalable layer

layer

<AVC and SVC elementary stream> set of VCL NAL units with the same values of dependency_id, quality_id and temporal_id, and the associated non-VCL NAL units

Note 1 to entry: The non-VCL NAL units as specified in ISO/IEC 14496-10.

Note 2 to entry: A scalable layer with any of dependency_id, quality_id, and temporal_id not equal to 0 enhances the video by one or more scalability levels in at least one direction (temporal, quality or spatial resolution).

Note 3 to entry: SVC uses a “layered” encoder design that results in a bitstream representing “coding layers”. In some publications, the “base layer” is the first quality layer of a specific coding layer. In some publications, the base layer is the scalable layer with the lowest priority. The SVC file format uses “scalable layer” or “layer” in a general way for describing nested bitstreams (using terms like AVC base layer or SVC enhancement layer).

3.1.26 scalable layer layer

<HEVC, layered HEVC and tiled HEVC elementary stream> set of VCL NAL units with the same values of `nuh_layer_id` and the associated non-VCL NAL units

Note 1 to entry: The non-VCL NAL units as specified in ISO/IEC 14496-10.

Note 2 to entry: A scalable layer with any of `dependency_id`, `quality_id`, and `temporal_id` not equal to 0 enhances the video by one or more scalability levels in at least one direction (temporal, quality or spatial resolution).

Note 3 to entry: SVC uses a “layered” encoder design that results in a bitstream representing “coding layers”. In some publications, the “base layer” is the first quality layer of a specific coding layer. In some publications, the base layer is the scalable layer with the lowest priority. The SVC file format uses “scalable layer” or “layer” in a general way for describing nested bitstreams (using terms like AVC base layer or SVC enhancement layer).

3.1.27 scalable layer representation

bitstream subset that is required for decoding the *scalable layer* (3.1.25), consisting of the scalable layer itself and all the scalable layers on which the scalable layer depends

Note 1 to entry: A scalable layer representation is also referred to as the representation of the scalable layer.

3.1.28 sub-picture

proper subset of coded slices of a layer representation

3.1.29 sub-picture tier

tier that consists of *sub-pictures* (3.1.28) IEC 14496-15:2017

Note 1 to entry: Any coded slice that is not included in the tier representation of a sub-picture tier is not to be referred to in inter prediction or inter-layer prediction for decoding of the sub-picture tier.

3.1.30 sub-layer

set of VCL NAL units with a particular value of `TemporalId` and the associated non-VCL NAL units

Note 1 to entry: The non-VCL NAL units are specified in ISO/IEC 23008-2.

3.1.31 SVC enhancement layer

layer that specifies a part of a scalable bitstream that enhances the video

Note 1 to entry: An SVC enhancement layer is represented by SVC VCL NAL units and the associated non-VCL NAL units and SEI messages.

Note 2 to entry: Usually, an SVC enhancement layer represents a spatial or coarse-grain scalability (CGS) coding layer (identified by a specific value of `dependency_id`).

3.1.32 SVC NAL unit

SVC VCL NAL unit and its associated non-VCL NAL units in an *SVC stream* (3.1.33)

Note 1 to entry: As specified in ISO/IEC 14496-10:2014, Annex G.

3.1.33

SVC stream

bitstream represented by the operating point for which `dependency_id` is equal to `mDid`, `temporal_id` is the greatest `temporal_id` value among `mOpSet`, and `quality_id` is the greatest `quality_id` value among `mOpSet`, where the greatest value of `dependency_id` of all the operating points represented by DTQ (`dependency_id`, `temporal_id` and `quality_id`) combinations is equal to `mDid`, and the set of all the operating points with `dependency_id` equal to `mDid` is `mOpSet`.

Note 1 to entry: The term “SVC stream” is referenced by “decoding/accessing the entire stream” in this document. There may be NAL units that are not required for decoding this operating point.

3.1.34

SVC VCL NAL unit

NAL unit with type 20, and NAL units with type 14 when the immediately following NAL units are AVC VCL NAL units

Note 1 to entry: As specified in ISO/IEC 14496-10:2014, Annex G.

Note 2 to entry: SVC VCL NAL units do not affect the decoding process of a legacy AVC decoder.

3.1.35

temporal layer representation

representation of a temporal layer

temporal layer and all lower temporal layers

3.1.36

tier

set of operating points within a track, providing information about the operating points and instructions on how to access the corresponding bitstream portions (using maps and groups)

Note 1 to entry: In SVC file format, a tier represents one or more scalable layers of an SVC bitstream. In the context of ISO/IEC 23008-2 video, the term tier is used to represent a part of the interoperability point representation consisting of profile, tier, and level. Readers should not be confused about these two different meanings of the word “tier”.

ISO/IEC 14496-15:2017

Note 2 to entry: The term “tier” is used in SVC file format to avoid confusion with the frequently used term layer. A tier represents a subset of a track and represents an operating point of an SVC bitstream. Tiers in a track subset the entire track, no matter whether the track references another track by extractors.

Note 3 to entry: An MVC or MVD tier represents a particular set of temporal subsets of a particular set of views.

3.1.37

tier representation

representation of the tier

bitstream subset that is required for decoding the tier, consisting of the tier itself and all the tiers on which the tier depends

3.1.38

video elementary stream

elementary stream containing access units made up of NAL units for coded picture data

3.1.39

video stream

self-contained independently decodable video bitstream

3.1.40

virtual base view

AVC compatible representation of an independently coded non-base view

Note 1 to entry: As specified in ISO/IEC 14496-10:2014, Annex H.