INTERNATIONAL STANDARD

ISO/IEC 14496-15

Third edition
2014-07-01

Information technology — Coding of audio-visual objects —

Part 15:
Carriage of network abstraction layer (NAL) unit structured video in 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

Reference number
ISO/IEC 14496-15:2014(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

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.

ISO/IEC 14496-15 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

This third edition cancels and replaces the second edition (ISO/IEC 14496-15:2010), which has been technically revised. It also incorporates the Amendment ISO/IEC 14496-15:2010/Amd.1:2011 and the Technical Corrigenda ISO/IEC 14496-15:2010/Cor.1:2011 and ISO/IEC 14496-15:2010/Cor.2:2012.

ISO/IEC 14496 consists of the following parts, under the general title *Information technology — Coding of audio-visual objects*:

- *Part 1: Systems*
- *Part 2: Visual*
- *Part 3: Audio*
- *Part 4: Conformance testing*
- *Part 5: Reference software*
- *Part 6: Delivery Multimedia Integration Framework (DMIF)*
- *Part 7: Optimized reference software for coding of audio-visual objects* [Technical Report]
- *Part 8: Carriage of ISO/IEC 14496 contents over IP networks*
- *Part 9: Reference hardware description* [Technical Report]
- *Part 10: Advanced Video Coding*

- *Part 11: Scene description and application engine*
- *Part 12: ISO base media file format*
- *Part 13: Intellectual Property Management and Protection (IPMP) extensions*
- *Part 14: MP4 file format*
- *Part 15: Carriage of network abstraction layer (NAL) unit structured video in the ISO base media file format*
- *Part 16: Animation Framework eXtension (AFX)*
- *Part 17: Streaming text format*
- *Part 18: Font compression and streaming*
- *Part 19: Synthesized texture stream*
- *Part 20: Lightweight Application Scene Representation (LASeR) and Simple Aggregation Format (SAF)*
- *Part 21: MPEG-J Graphics Framework eXtension (GFX)*
- *Part 22: Open Font Format*
- *Part 23: Symbolic Music Representation*
- *Part 24: Audio and systems interaction*
- *Part 25: 3D Graphics Compression Model*
- *Part 26: Audio conformance*
- *Part 27: 3D Graphics conformance*
- *Part 28: Composite font representation*

Introduction

This part of ISO/IEC 14496 defines a storage format based on, and compatible with, the ISO Base Media File Format (ISO/IEC 14496-12 and ISO/IEC 15444-12), 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 part of ISO/IEC 14496 enables video streams formatted as Network Adaptation Layer Units (NAL Units) to

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

This part of ISO/IEC 14496 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) and 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. Information may be obtained from the companies listed in Annex F.

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

INTERNATIONAL STANDARD **ISO/IEC 14496-15:2014(E)**

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 part of ISO/IEC 14496 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, in whole or in part, are normatively referenced in this document and are indispensable for its application. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

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

ISO/IEC 14496-12, *Information technology — Coding of audio-visual objects — Part 12: ISO base media file format*[1]

ISO/IEC 23008-2, *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.

3.1.1
aggregator
in-stream structure using a NAL unit header

NOTE Aggregators are used to group NAL units belonging to the same sample.

[1] ISO/IEC 14496-12 is technically identical to ISO/IEC 15444-12.

3.1.2
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 Annex G or Annex H)

NOTE 1 The AVC base layer is represented by AVC VCL NAL units and associated non-VCL NAL units.

NOTE 2 The AVC base layer itself can be a temporal scalable bitstream.

3.1.3
AVC NAL unit

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

3.1.4
AVC VCL NAL unit

NAL unit with type 1 to 5 (inclusive) as specified in ISO/IEC 14496-10

3.1.5
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 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.

3.1.6
extractor

in-stream structure using a NAL unit header including a NAL unit header extension

NOTE Extractors contain instructions on how to extract data from other tracks. Logically an Extractor can be seen as a ‘link’. While accessing a track containing Extractors, the Extractor is replaced by the data it is referencing.

3.1.7
in-stream structure

structure residing within sample data

3.1.8
MVC VCL NAL unit

NAL unit with type 20, and NAL units with type 14, as specified in ISO/IEC 14496-10, when the immediately following NAL units are AVC VCL NAL units.

NOTE MVC VCL NAL units do not affect the decoding process of a legacy AVC decoder.

3.1.9
operating point

subset of a scalable bitstream, representing in SVC a particular spatial resolution, temporal resolution, and quality, or in MVC a set of target output views

NOTE 1 Each operating point consists of all the data needed to decode this particular bitstream subset.

NOTE 2 In an SVC stream an operating point 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 3 The bitstream subset of an MVC 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.

NOTE 4 An operating point is referred to as an operation point in Annex H of ISO/IEC 14496-10 or in ISO/IEC 23008-2.

3.1.10
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 This term is used to refer to all types of parameter sets.

3.1.11
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

3.1.12
prefix NAL unit

NAL units with type 14 as specified in ISO/IEC 14496-10

NOTE 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.

3.1.13
scalable layer; layer

set of VCL NAL units with the same values of dependency_id, quality_id, and temporal_id, and the associated non-VCL NAL units as specified in ISO/IEC 14496-10.

NOTE 1 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 2 SVC uses a “layered” encoder design which 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).

3.1.14
scalable layer representation

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

NOTE A scalable layer representation is also referred to as the representation of the scalable layer.

3.1.15
sub-picture

proper subset of coded slices of a layer representation

3.1.16
sub-picture tier

tier that consists of sub-pictures

NOTE 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.

3.1.17
SVC enhancement layer

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

NOTE 1 An SVC enhancement layer is represented by SVC VCL NAL units and the associated non-VCL NAL units and SEI messages.

NOTE 2 Usually an SVC enhancement layer represents a spatial or coarse-grain scalability (CGS) coding layer (identified by a specific value of dependency_id).

3.1.18
SVC NAL unit

SVC VCL NAL unit and its associated non-VCL NAL units in an SVC stream

3.1.19
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 The term "SVC stream" is referenced by 'decoding/accessing the entire stream' in this document. There may be NAL units which are not required for decoding this operating point.

3.1.20
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 SVC VCL NAL units do not affect the decoding process of a legacy AVC decoder.

3.1.21
temporal layer representation
representation of a temporal layer

temporal layer and all lower temporal layers

3.1.22
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 A tier represents one or more scalable layers of an SVC bitstream.

NOTE 2 The term "tier" is used 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 An MVC tier represents a particular set of temporal subsets of a particular set of views.

3.1.23
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

3.1.24
video elementary stream

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

3.1.25
virtual base view

AVC compatible representation of an independently coded non-base view

NOTE The virtual base view of an independently coded non-base view is created according to the process specified in H.8.5.5 of ISO/IEC 14496-10. Samples containing data units of an independently coded non-base view and samples of the virtual base view are aligned by decoding times.

3.2 Abbreviated terms

AVC	Advanced Video Coding. Where contrasted with SVC or MVC in this International Standard, this term refers to the main part of ISO/IEC 14496-10, including neither Annex G (Scalable Video Coding) nor Annex H (Multiview Video Coding)
BLA	Broken Link Access
CRA	Clean Random Access
CTU	Coding Tree Unit
HEVC	High Efficiency Video Coding
FF	File Format
HRD	Hypothetical Reference Decoder
IDR	Instantaneous Decoding Refresh
MVC	MultiviewVideo Coding [refers to ISO/IEC 14496-10 when the techniques in Annex H (Multiview Video Coding) are in use]
NAL	Network Abstraction Layer
PPS	Picture Parameter Set
ROI	Region-Of-Interest
SEI	Supplementary Enhancement Information
SPS	Sequence Parameter Set
STSA	Step-wise Temporal Sub-layer Access
SVC	Scalable Video Coding [refers to ISO/IEC 14496-10 when the techniques in Annex G (Scalable Video Coding) are in use]
TSA	Temporal Sub-layer Access
VCL	Video Coding Layer
VPS	Video Parameter Set

4 General Definitions

4.1 Introduction

The specifications in this clause apply to all coding systems identified by chapters in this specification, unless specifically over-ridden by definitions in the clause for a specific coding system.

The following table summarizes the correspondences between the sets of terminology used in video specifications and the ISO Base Media File Format.

Table 1 – Correspondence of terms in video and ISO Base Media File Format

Video	ISO Base Media File Format
-	Movie
Bitstream	Track
Access Unit	Sample

4.2 Elementary stream structure

This specification concerns video coding systems that specify a set of Network Abstraction Layer (NAL) units, which contain different types of data. This subclause specifies the format of the elementary streams for storing such content.

4.3 Sample and Configuration definition

4.3.1 Introduction

Sample: A sample is an access unit as defined in the appropriate specification.

Parameter set sample: A parameter set sample is a sample in a parameter set stream which shall consist of those parameter set NAL units that are to be considered as if present in the video elementary stream at the same instant in time.

4.3.2 Canonical order and restrictions

The elementary stream is stored in the ISO Base Media File Format in a *canonical* format. The canonical format is as *neutral* as possible so that systems that need to customize the stream for delivery over different transport protocols — MPEG-2 Systems, RTP, and so on — should not have to *remove* information from the stream while being free to *add* to the stream. Furthermore, a canonical format allows such operations to be performed against a known initial state.

The canonical stream format is an elementary stream that satisfies the following conditions:

- **Video data NAL units**: All video data NAL units for a single picture shall be contained with the sample whose decoding time and composition time are those of the picture. Each sample shall contain at least one video data NAL unit of the primary picture.

- **SEI NAL units**: All SEI NAL units shall be contained in the parameter set arrays, or in the sample whose decoding time is at the time, or immediately precedes the time (with no intervening samples), when the SEI messages come into effect instantaneously. In general, SEI messages for a picture shall be included in the sample containing that picture and that SEI messages pertaining to a sequence of pictures shall be included in the sample containing the first picture of the sequence to which the SEI message pertains. The order of SEI messages within a sample is as defined in the applicable video coding standard.

- The sequence of NAL units in an elementary stream and within a single sample must be in a valid decoding order for those NAL units as specified in the applicable video coding standard.

- **All timing information is external to stream.** Picture Timing SEI messages that define presentation or composition timestamps may be included in the video elementary stream, as these messages contain other information than timing, and may be required for conformance checking. However, all timing information is provided by the information stored in the various sample metadata tables, and this information over-rides any timing provided in the video layer. Timing provided within the video stream in this file format should be ignored as it may contradict the timing provided by the file format and may not be correct or consistent within itself.

  NOTE This constraint is imposed due to the fact that post-compression editing, combination, or re-timing of a stream at the file format level may invalidate or make inconsistent any embedded timing information present within the video stream.

- **No start codes.** The elementary streams shall not include start codes. As stored, each NAL unit is preceded by a length field as specified in 4.3.3; this enables easy scanning of the sample's NAL units. Systems that wish to deliver, from this file format, a stream using start codes will need to reformat the stream to insert those start codes.

4.3.3 Sample format

4.3.3.1 Definition

This subclause defines the structure of the samples. Samples are externally framed and have a size supplied by that external framing. The syntax of a sample is configured via the decoder specific configuration for the elementary stream. An example of the structure of a video sample is depicted in the following figure.

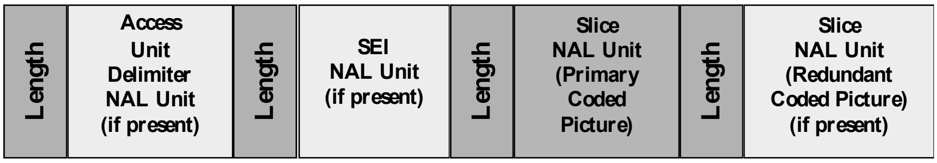
Length
Access Unit Delimiter NAL Unit (if present)
Length
SEI NAL Unit (if present)
Length
Slice NAL Unit (Primary Coded Picture)
Length
Slice NAL Unit (Redundant Coded Picture) (if present)

Figure 1 — Example structure of a sample

An access unit is made up of a set of NAL units. Each NAL unit is represented with a:

- *Length*: Indicates the length in bytes of the following NAL unit. The length field can be configured to be of 1, 2, or 4 bytes.

- *NAL Unit*: Contains the NAL unit data as specified in the applicable video coding standard.