INTERNATIONAL STANDARD

ISO/IEC 14496-15

First edition
2004-04-15

AMENDMENT 2
2008-07-15

Information technology — Coding of audio-visual objects —

Part 15: Advanced Video Coding (AVC) file format

AMENDMENT 2: File format support for Scalable Video Coding (SVC)

Technologies de l'information — Codage des objets audiovisuels —

Partie 15: Format de fichier de codage vidéo avancé (AVC)

AMENDEMENT 2: Support de format de fichier pour codage vidéo extensible (SVC)

Reference number
ISO/IEC 14496-15:2004/Amd.2:2008(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

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.

Amendment 2 to ISO/IEC 14496-15:2004 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information Technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*

Information technology — Coding of audio-visual objects —

Part 15:
Advanced Video Coding (AVC) file format

AMENDMENT 2: File format support for Scalable Video Coding (SVC)

Add to the Introduction:

This International Standard defines the storage for both plain AVC and SVC video streams, where 'plain AVC' refers to the main part of ISO/IEC 14496-10, not including Annex G (Scalable Video Coding), and SVC refers to ISO/IEC 14496-10 when the techniques in Annex G (Scalable Video Coding) are in use. Specific techniques are introduced for the handling of scalable streams, enabling their use, and assisting the extraction of subsets of scalable streams.

Add the following to the end of Clause 1 (Scope):

The file format for storage of SVC content, as defined in Annexes A-E, uses the existing capabilities of the ISO base media file format and the AVC file format. In addition, the following new extensions to support SVC-specific features are specified:

- **Scalable Grouping**: A structuring and grouping mechanism to indicate the association of NAL units with different types and hierarchy levels of scalability.
- **Aggregator**: A structure to enable efficient scalable grouping of NAL units by changing irregular patterns of NAL units into regular patterns of aggregated data units.
- **Extractor**: A structure to enable efficient extraction of NAL units from other tracks than the one containing the media data.
- **Temporal metadata statements**: Structures for storing time-aligned information of media samples.
- **AVC Compatibility**: A provision for storing an SVC bitstream in an AVC compatible manner, such that the AVC compatible base layer can be used by any existing AVC file format compliant reader.

Add the following terms (or replace when the item exists) to 3.2 maintaining alphabetical order:

AVC	Advanced Video Coding. Where contrasted with SVC in this International Standard, this term refers to the main part of ISO/IEC 14496-10, not including Annex G (Scalable Video Coding)
FF	File Format
ROI	Region-Of-Interest
SVC	Scalable Video Coding. Refers to ISO/IEC 14496-10 when the techniques in Annex G (Scalable Video Coding) are in use
VCL	Video Coding Layer

Replace the text of 4.1 with the following:

The technologies originally documented in Clause 4 are now defined in ISO/IEC 14496-12:2008 (technically identical to ISO/IEC 15444-12:2008).

Replace the text of 4.2 (file identification) with the following:

See 6.3 in ISO/IEC 14496-12:2008.

Replace the text of 4.3 (independent and disposable samples) with the following:

See 8.6.4 in ISO/IEC 14496-12:2008 for the definition of this box.

Replace the text of 4.4 (sample groups) with the following:

See 8.9 in ISO/IEC 14496-12:2008.

Replace the text of 4.5 (random access recovery points) with the following:

See 10.1 in ISO/IEC 14496-12:2008.

Replace the text of 4.6 (representation of new structures) with the following:

See 8.9.4 in ISO/IEC 14496-12:2008.

Add to the end of 5.3.3 (Track Structure):

A video stream is represented by one or more video tracks in a file.

If there is more than one track representing scalable aspects of a single stream, then they form alternatives to each other, and the field '`alternate_group`' should be used, or the composition system used should select one of them, as appropriate. See 10.3 "Grouping and labelling of tracks" of ISO/IEC 14496-12:2008 for informative labelling of why tracks are members of alternate groups.

Add this new subclause as 5.3.16 Definition of a sub-sample for AVC:

For the use of the sub-sample information box (7.7 of ISO/IEC 14496-12:2008) in an AVC stream, a sub-sample is defined as one or more contiguous NAL units within a sample and having the same value of the following fields; `RefPicFlag`, `RedPicFlag` and `VclNalUnitFlag`. Each sub-sample includes both NAL unit(s) and their preceding NAL unit length field(s). The presence of this box is optional; however, if present in a track containing AVC data, it shall have the semantics defined here.

The subsample_priority field shall be set to a value in accordance with the specification of this field in ISO/IEC 14496-12.

The discardable field shall be set to 1 only if this sample can still be decoded if this sub-sample is discarded (e.g. the sub-sample consists of an SEI NAL unit, or a redundant coded picture).

The reserved field is defined for AVC as follows:

```
unsigned int(1) RefPicFlag;
unsigned int(1) RedPicFlag;
unsigned int(1) VclNalUnitFlag;
unsigned int(29) reserved = 0;
```

`RefPicFlag` equal to 0 indicates that all the NAL units in the sub-sample have `nal_ref_idc` equal to 0. `RefPicFlag` equal to 1 indicates that all the NAL units in the sub-sample have `nal_ref_idc` greater than 0.

`RedPicFlag` equal to 0 indicates that all the NAL units in the sub-sample have `redundant_pic_cnt` equal to 0. `RedPicFlag` equal to 1 indicates that all the NAL units in the sub-sample have `redundant_pic_cnt` greater than 0.

`VclNalUnitFlag` equal to 0 indicates that all NAL units in the sub-sample are non-VCL NAL units. Value 1 indicates that all NAL units in the sub-sample are VCL NAL units.

Insert the following Annexes A to E:

Annex A
(normative)

SVC elementary stream and sample definitions

A.1 Introduction

This annex specifies the basic storage format of SVC data. It extends the definitions of the storage format of AVC in Clause 5.

A.2 Terms and definitions

For the purpose of Annex A to Annex E, inclusive, the following terms and definitions apply. The definitions in ISO/IEC 14496-10 (including Annex G) also apply.

A.2.1
Aggregator
Aggregators are in-stream structures using a NAL unit header including a NAL unit header SVC extension, with a NAL unit type equal to 30. Aggregators are used to group NAL units belonging to the same sample.

A.2.2
AVC base layer
The AVC base layer is the maximum subset of a scalable bitstream that is AVC compatible – a bitstream not using any of the functionality of ISO/IEC 14496-10 Annex G. The AVC base layer is represented by AVC VCL NAL units and associated non-VCL NAL units.

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

A.2.3
AVC NAL units
AVC VCL NALs collectively refer to AVC VCL NAL units and their associated non-VCL NAL units in a bitstream.

A.2.4
AVC VCL NAL unit
AVC VCL NAL units are NAL units with type in the range of 1 to 5 (inclusive).

A.2.5
Extraction path
An extraction path is a 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.

A.2.6
Extractor
Extractors are in-stream structures using a NAL unit header including a NAL unit header SVC extension, with a NAL unit type equal to 31. 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.

A.2.7
In-stream
In-stream structures reside within sample data.

A.2.8
Operating point
A subset of a scalable bitstream, representing a particular spatial resolution, temporal resolution, and quality. Each operating point consists of all the data needed to decode this particular bitstream subset.

NOTE 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.

A.2.9
Prefix NAL unit
Prefix NAL units are NAL units with type 14. 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.

A.2.10
Scalable layer
A scalable layer consists of a set of VCL NAL units with the same values of dependency_id, quality_id, and temporal_id and the associated non-VCL NAL units. 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). A scalable layer may also be simply referred to a layer.

NOTE 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).

A.2.11
Scalable layer representation
A scalable layer representation refers to the bitstream subset that is required for decoding the scalable layer, and consists of the scalable layer itself and all the scalable layers on which the scalable layer depends. A scalable layer representation is also referred to as the representation of the scalable layer.

A.2.12
Sub-picture
A sub-picture consists of a proper subset of coded slices of a layer representation.

A.2.13
Sub-picture tier
A sub-picture tier is a tier that consists of sub-pictures. Any coded slice that is not included in the tier representation of a sub-picture tier shall not be referred to in inter prediction or inter-layer prediction for decoding of the sub-picture tier.

A.2.14
SVC enhancement layer
An SVC enhancement layer specifies a part of a scalable bitstream that enhances the video. An SVC enhancement layer is represented by SVC VCL NAL units and the associated non-VCL NAL units and SEI messages.

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

A.2.15
SVC NAL units
SVC NAL units collectively refer to SVC VCL NAL units and their associated non-VCL NAL units in an SVC stream.

A.2.16
SVC stream
Let the greatest value of dependency_id of all the operating points represented by DTQ (dependency_id, temporal_id and quality_id) combinations be equal to mDid, and the set of all the operating points with dependency_id equal to mDid be mOpSet. SVC stream refers to the 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. 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.

A.2.17
SVC VCL NAL unit
SVC VCL NAL units are NAL units with type 20, and NAL units with type 14 when the immediately following NAL units are AVC VCL NAL units. SVC VCL 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 SVC VCL NAL units is undefined.

A.2.18
Tier
Tiers define a 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). A tier represents one or more scalable layers of an SVC bitstream.

NOTE 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.

A.2.19
Tier representation
A tier representation refers to the bitstream subset that is required for decoding the tier, and consists of the tier itself and all the tiers on which the tier depends. A tier representation is also referred to as the representation of the tier.

A.3 Elementary stream structure

SVC streams are stored in accordance with subclause 5.1, with the following definition of an SVC video elementary stream:

- **SVC Video Elementary Streams** shall contain all video coding related NAL units (i.e. those NAL units containing video data or signalling video structure) and may contain non-video coding related NAL units such as SEI messages and access unit delimiter NAL units. Also Aggregators (see B.2) or Extractors (see B.3) may be present. Aggregators and Extractors shall be processed as defined in this International Standard (e.g. shall not directly be placed in the output buffer while accessing the file). Other NAL units that are not expressly prohibited may be present, and if they are unrecognized they should be ignored (e.g. not placed in the output buffer while accessing the file).

SVC streams may also be stored using associated parameter set streams, if needed.

For SVC streams, Table 1 is updated as follows; only entries where the definition for SVC differs from AVC, are shown.

Table A.1 — NAL Unit Types in SVC and AVC Streams

Value of nal_unit_type	Description	AVC video elementary stream	SVC video elementary stream	Parameter set elementary stream
14	Prefix NAL unit in scalable extension prefix_nal_unit_rbsp( )	Not specified	Yes	No
15	Subset sequence parameter set subset_seq_parameter_set_rbsp( )	Not specified	No. If parameter set elementary stream is not used, Subset SPS shall be stored in the Decoder Specific Information.	Yes
20	Coded slice in scalable extension slice_layer_in_scalable_extension_rbsp( )	Not specified	Yes	No
24 – 29	Not specified	Not specified	Not specified	Not specified
30	Aggregator	Not specified	Yes	No
31	Extractor	Not specified	Yes	No

There may be AVC VCL NAL units, SVC VCL NAL units and other NAL units, i.e. non-VCL NAL units, present in an SVC video elementary stream. Additionally, there may be Aggregator NAL units and Extractor NAL units present in an SVC video elementary stream.

An AVC VCL NAL unit in an SVC video elementary stream conforming to one or more profiles specified in Annex G of ISO/IEC 14496-10 shall be immediately preceded by a prefix NAL unit containing the scalability information for the AVC VCL NAL unit. In this file format an AVC VCL NAL unit and the immediately preceding prefix NAL unit are logically seen as one NAL unit: the prefix NAL unit provides the scalability information and the AVC VCL NAL unit provides the NAL unit type and payload.

A.4 Use of the AVC file format

The SVC file format is an extension of the AVC file format defined in this International Standard.

Subclause 5.3.12 is defined for use with plain AVC streams. Its use with SVC streams is deprecated.

A.5 Sample and configuration definition

A.5.1 Introduction

SVC Sample: An SVC sample is also an access unit as defined in subclause 7.4.1.2 of ISO/IEC 14496-10.

A.5.2 Canonical order and restrictions

A.5.2.1 Restrictions

The following restrictions apply to SVC data in addition to the requirements in subclause 5.2.2:

- **SVC coded slice NAL units** (Coded slices in scalable extension): All SVC coded slice NAL units for a single instant in time shall be contained in the sample whose composition time is that of the picture represented by the access unit. An SVC sample shall contain at least one AVC or SVC VCL NAL unit.

- **Prefix NAL units** (Prefix NAL unit in scalable extension): Each prefix NAL unit is placed immediately before the corresponding AVC VCL NAL unit, providing scalability information about the AVC VCL NAL unit.

  NOTE Prefix NAL units may also be associated with filler data NAL units, which are not present in elementary streams.

- **Aggregators/Extractors**: The order of all NAL units included in an Aggregator or referenced by an Extractor is exactly the decoding order as if these NAL units were present in a 'plain' sample. After processing the Aggregator or the Extractor, all NAL units must be in valid decoding order as specified in ISO/IEC 14496-10.

A.5.2.2 Decoder configuration record

When the decoder configuration record defined in subclause 5.2.4.1 is used for a stream which can be interpreted as either an SVC or AVC stream, the AVC decoder configuration record shall reflect the properties of the AVC compatible base layer, e.g. it shall contain only parameter sets needed for decoding the AVC base layer.

A parameter set stream may be used with SVC streams, as with AVC streams. In that case, parameter sets shall not be included in the decoder configuration record.

Sequence parameter sets are numbered in order of storage from 1 to `numOfSequenceParameterSets` or `numOfPictureParameterSets` respectively. Sequence and Picture parameter sets stored in this record in a file may be referenced using this 1-based index by the `InitialParameterSetBox`.

The `SVCDecoderConfigurationRecord` is structurally identical to an `AVCDecoderConfigurationRecord`. However, the reserved bits preceding and succeeding the `lengthSizeMinusOne` field are re-defined. The syntax is as follows:

```
aligned(8) class SVCDecoderConfigurationRecord {
   unsigned int(8) configurationVersion = 1;
   unsigned int(8) AVCProfileIndication;
   unsigned int(8) profile_compatibility;
   unsigned int(8) AVCLevelIndication;
   bit(1) complete_represenation;
   bit(5) reserved = '11111'b;
   unsigned int(2) lengthSizeMinusOne;
   bit(1) reserved = '0'b;
   unsigned int(7) numOfSequenceParameterSets;
   for (i=0; i< numOfSequenceParameterSets; i++) {
      unsigned int(16) sequenceParameterSetLength ;
      bit(8*sequenceParameterSetLength) sequenceParameterSetNALUnit;
   }
   unsigned int(8) numOfPictureParameterSets;
   for (i=0; i< numOfPictureParameterSets; i++) {
      unsigned int(16) pictureParameterSetLength;
      bit(8*pictureParameterSetLength) pictureParameterSetNALUnit;
   }
}
```

The semantics of the fields `AVCProfileIndication`, `profile_compatibility`, and `AVCLevelIndication` differ from the `AVCDecoderConfigurationRecord` as follows:

The fields `AVCProfileIndication`, `AVCLevelIndication` carry the profile and level indications, respectively, indicating the profile and level of the entire scalable stream in this track. They, and the `profile_compatibility` field, must have values such that a conforming SVC decoder is able to decode bitstreams conforming to the profile, level and profile compatibility flags indicated in any of the sequence parameter sets or subset sequence parameter sets contained in this record.

The semantics of other fields are as follows, or are as defined for an `AVCDecoderConfigurationRecord`:

`complete_representation` is set on a minimal set of tracks that contain a portion of the original encoded scalable stream, as defined in A.6.1. Other tracks may be removed from the file without loss of any portion of the original encoded bitstream, and, once the set of tracks has been reduced to only those in the complete subset, any further removal of a track removes a portion of the encoded information.

`numOfSequenceParameterSets` indicates the number of SPSs and subset SPSs that are used for decoding the SVC elementary stream. The value of `numOfSequenceParameterSets` shall be in the range of 0 to 64, inclusive.

`SequenceParameterSetLength` indicates the length in bytes of the SPS or subset SPS NAL unit.

`SequenceParameterSetNALUnit` contains a SPS or subset SPS NAL unit. SPSs shall occur in order of ascending parameter set identifier with gaps being allowed. Subset SPSs shall occur in order of ascending parameter set identifier with gaps being allowed. Any SPS shall occur before all the subset SPSs, if any.

A.6 Derivation from the ISO base media file format

A.6.1 SVC track structure

A scalable video stream is represented by one or more video tracks in a file. Each track represents one or more operating points of the scalable stream. A scalable stream may, of course, be further thinned, if desired.

There is a minimal set of one or more tracks that, when taken together, contain the complete set of encoded information. All these tracks shall have the flag "`complete_representation`" set in all their sample entries. This group of tracks that form the complete encoded information are called the "complete subset".

Let the lowest operating point be the one of all the operating points represented by DTQ (dependency_id, temporal_id and quality_id) combinations that has the least values of dependency_id, temporal_id and quality_id, respectively. The track that has the flag "`complete_representation`" set and contains the lowest operating point shall be nominated as the 'scalable base track'. All the other tracks that are part of the same scalable encoded information shall be linked to this base track by means of a track reference of type '`sbas`' (scalable base). The complete encoded information can be retained when the tracks included in the "complete subset" are retained; all other tracks shall be extractions, subsets, copies or re-orderings of the complete subset.

NOTE 1 An alternate group may also include completely independent bitstreams, as well as alternative operating points of the same bitstream. The SVC tracks in the alternate group must be examined to see how many scalable base tracks are identified.

NOTE 2 "A scalable bitstream" may require more than one track to represent it (consider a stream with a low-resolution, low-frame-rate base layer, and a high resolution enhancement layer, and a high frame-rate enhancement layer, but missing the data for high resolution high frame-rate). However, such a scalable bitstream is typically a non-conforming bitstream.

All the tracks sharing the same scalable base track must share the same timescale.

A.6.2 Data sharing and extraction

Different tracks may logically share data. This sharing can take one of the following two forms:

a) The sample data is copied from one track into another track (and possibly compacted or re-interleaved with other data, such as audio). This creates larger overall files, but the low bit rate data may be compacted and/or interleaved with other material, for ease of extraction.

b) There may be instructions on how to perform this copy at the time that the file is read.

For the second case, Extractors (defined in B.3) are used.

A.6.3 SVC video stream definition

A.6.3.1 Sample description name and format

A.6.3.1.1 Definition

Types: 'avc2', 'avcC', 'svc1','svcC','seib'
Container: Sample Table Box ('stbl')
Mandatory: Either the avc1, or avc2 or svc1 box is mandatory.
Quantity: One or more sample entries may be present

If an SVC elementary stream contains a usable AVC compatible base layer, then an AVC visual sample entry ('avc1' or 'avc2') shall be used. Here, the entry shall contain initially an AVC Configuration Box, possibly followed by an SVC Configuration Box as defined below. The AVC Configuration Box documents the Profile, Level and Parameter Set information pertaining to the AVC compatible base layer as defined by the `AVCDecoderConfigurationRecord`. The SVC Configuration Box documents the Profile, Level and Parameter Set information pertaining to the entire stream containing the SVC compatible enhancement layers as defined by the `SVCDecoderConfigurationRecord`, stored in the SVCConfigurationBox.

For all sample entries except for 'svc1', i.e. 'avc1' and 'avc2', the width and height fields in the sample entry document the AVC base layer. For an 'svc1' sample entry, the width and height document the resolution achieved by decoding the entire stream.

If the SVC elementary stream does not contain a usable AVC base layer, then an SVC visual sample entry ('svc1') shall be used. The SVC visual sample entry shall contain an SVC Configuration Box, as defined below. This includes an `SVCDecoderConfigurationRecord`, as defined in this International Standard.

The `lengthSizeMinusOne` field in the SVC and AVC configurations in any given sample entry shall have the same value.

A priority assignment URI provides the name (in the URI space) of a method used to assign priority_id values. When it occurs in an AVC or SVC sample entry, exactly one URI shall be present, that documents the priority_id assignments in the stream. The URI is treated here as a name only; it should be de-referenceable, though this is not required. File readers may be able to recognize some methods and thereby know what stream extraction operations based on priority_id would do.

The sample entry name 'avc1' may only be used when the entire stream is a compliant and usable AVC stream as viewed by an AVC decoder operating under the configuration (including profile and level) given in the AVCConfigurationBox. The file format specific structures that resemble NAL units may be present but must not be used to access the AVC base data; that is, the AVC data must not be contained in Aggregators (though they may be included within the bytes referenced by the additional_bytes field) nor referenced by Extractors. The sample entry name 'avc2' indicates that, in order to form the intended AVC stream, Extractors must be replaced with the data they are referencing, and Aggregators must be examined for contained NAL Units. Extractors or aggregators may be used for SVC VCL NAL units in 'avc1', 'avc2' or 'svc1' tracks.

NOTE When AVC compatibility is indicated, it may be necessary to indicate an unrealistic level for the AVC base layer, to accommodate the bit rate of the entire stream, because all the NAL units are considered as included in the AVC base layer and hence may be fed to the decoder, which is expected to discard those NAL unit it does not recognize. This case happens when the 'avc1' sample entry is used and both AVC and SVC configurations are present.

